170010

STIC-Biotech/ChemLib

From: Walicka, Malgorzata
Sent: Saturday, October 29, 2005 10:56 AM
To: STIC-Biotech/ChemLib

Please search SEQ ID NO:1 and 16 in 09/774,954. Please use also pending data.

Thank you.

Malgorzata A. Walicka, Ph.D.
Patent Examiner
Art Unit 1652, Recombinant Enzymes
USPTO, Remsen Building, Room 2C76
400 Dulany St.
Alexandria, VA 22313
Mail Room 2C70
Tel. (571) 272-0944, fax (571) 273-0944

****
Searcher:_____
Searcher Phone:_____
Date Searcher Picked up:11/1/05
Date completed:11/9/05
Searcher Prep Time:_____
Online Time:_____

****
Type of Search
NA#3 AA#:_____
S/L:_____ Oligomer:_____
Encode/Transl:_____
Structure #:_____Text:____
Inventor:_____ Litigation:___

****
Vendors and cost where applicable
STN:_____
DIALOG:_____
QUESTEL/ORBIT:_____
LEXIS/NEXIS:_____
SEQUENCE SYSTEM:
WWW/Internet:_____
Other (Specify):_____

GenCore version 5.1.6

OM nucleic - nucleic search, using sw model

Run on: November 8, 2005, 23:53:44 ; Search time 6830.38 Seconds
(without alignments)
10740.434 Million cell updates/sec

Title: US-09-774-954-1
Perfect score: 1514
Sequence: 1 ATGCCCGCGGGCTCCTGGGA.....TTTTGAAAATCATTGAATTC 1514

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 4708233 seqs, 24227607955 residues

Total number of hits satisfying chosen parameters: 9416466

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl:*
1: gb_ba:*
2: gb_htg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vi:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	1514	100.0	1514	6	AR163451	AR163451 Sequence
2	1514	100.0	1514	6	BD103188	BD103188 O-fucosyl
3	1512.4	99.9	5249	9	AF375884	AF375884 Homo sapi
4	1512.4	99.9	11284	6	BD103190	BD103190 O-fucosyl
5	1510.8	99.8	5189	9	D80002	D80002 Homo sapien
6	1509.8	99.7	5266	6	AX780153	AX780153 Sequence
7	1507.6	99.6	4560	9	HSM803330	AL832023 Homo sapi
8	1447.4	95.6	5009	6	AR163453	AR163453 Sequence
9	1447.4	95.6	5009	6	BD103191	BD103191 O-fucosyl
10	1163.4	76.8	1300	6	AR163452	AR163452 Sequence
11	1163.4	76.8	1300	6	BD103189	BD103189 O-fucosyl
12	1100	72.7	1100	6	AR163457	AR163457 Sequence
13	1100	72.7	1100	6	BD103196	BD103196 O-fucosyl
14	1096.4	72.4	1167	6	CQ727777	CQ727777 Sequence
15	1086.8	71.8	1167	9	AJ781500	AJ781500 Pan trogl
16	916.2	60.5	1176	4	AY344580	AY344580 Bos tauru
17	908.2	60.0	1176	4	AJ781503	AJ781503 Bos tauru
18	901.6	59.6	1134	4	SSC567917	AJ567917 Sus scrof
19	889.6	58.8	1188	10	AJ781499	AJ781499 Rattus no
20	889.6	58.8	1302	10	AF375885	AF375885 Mus muscu
21	889.6	58.8	1529	10	BC046295	BC046295 Mus muscu
22	752.8	49.7	1056	4	AY344581	AY344581 Bos tauru
23	652.2	43.1	1732	5	GGA535754	AJ535754 Gallus ga
24	652.2	43.1	2546	5	AJ720352	AJ720352 Gallus ga
25	608.6	40.2	145414	9	HSA392M18	AL121897 Human DNA
26	586.4	38.7	1502	5	BC082519	BC082519 Xenopus t
27	586.4	38.7	1538	5	AJ781498	AJ781498 Silurana
28	563.6	37.2	1155	5	AJ606070	AJ606070 Fugu rubr
29	552.8	36.5	1505	5	XLA514425	AJ514425 Xenopus l
30	539	35.6	1164	5	AJ781504	AJ781504 Tetraodon
31	538.4	35.6	1549	5	DRE579536	AJ579536 Danio rer
32	471.4	31.1	1524	9	BC000582	BC000582 Homo sapi
33	445.4	29.4	920	5	AJ719585	AJ719585 Gallus ga
34	201.6	13.3	1841	3	AK112708	AK112708 Ciona int
35	193.4	12.8	2393	9	AK124093	AK124093 Homo sapi
c 36	188.8	12.5	240336	2	AC094844	AC094844 Rattus no
c 37	186.6	12.3	240510	2	AC134371	AC134371 Rattus no
38	184.6	12.2	200329	10	AC078911	AC078911 Mus muscu
39	184.6	12.2	208614	10	AL807380	AL807380 Mouse DNA
40	184	12.2	1146	3	AJ781502	AJ781502 Ciona sav
41	172.8	11.4	1209	3	AJ831490	AJ831490 Drosophil
42	168	11.1	1295	3	AY118651	AY118651 Drosophil
43	168	11.1	1320	6	AR509393	AR509393 Sequence
44	166.8	11.0	1209	3	AJ781501	AJ781501 Drosophil
45	166.4	11.0	1209	6	CQ590403	CQ590403 Sequence

ALIGNMENTS

RESULT 1
AR163451
LOCUS AR163451 1514 bp DNA linear PAT 17-OCT-2001
DEFINITION Sequence 2 from patent US 6270987.
ACCESSION AR163451
VERSION AR163451.1 GI:16234056
KEYWORDS .
SOURCE Unknown.
ORGANISM Unknown.
Unclassified.
REFERENCE 1 (bases 1 to 1514)
AUTHORS Wang,Y. and Spellman,M.W.
TITLE O-fucosyltransferase
JOURNAL Patent: US 6270987-A 2 07-AUG-2001;
FEATURES Location/Qualifiers
source 1. .1514
/organism="unknown"
/mol_type="unassigned DNA"
ORIGIN

Query Match 100.0%; Score 1514; DB 6; Length 1514;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1514; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 ATGCCCGCGGGCTCCTGGGACCCGGCCGGTTACCTGCTCTACTGCCCCTGCATGGGGCGC 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGCCCGCGGGCTCCTGGGACCCGGCCGGTTACCTGCTCTACTGCCCCTGCATGGGGCGC 60

Qy         61 TTTGGGAACCAGGCCGATCACTTCTTGGGCTCTCTGGCATTTGCAAAGCTGCTAAACCGT 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 TTTGGGAACCAGGCCGATCACTTCTTGGGCTCTCTGGCATTTGCAAAGCTGCTAAACCGT 120

Qy        121 ACCTTGGCTGTCCCTCCTTGGATTGAGTACCAGCATCACAAGCCTCCTTTCACCAACCTC 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 ACCTTGGCTGTCCCTCCTTGGATTGAGTACCAGCATCACAAGCCTCCTTTCACCAACCTC 180

Qy        181 CATGTGTCCTACCAGAAGTACTTCAAGCTGGAGCCCCTCCAGGCTTACCATCGGGTCATC 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 CATGTGTCCTACCAGAAGTACTTCAAGCTGGAGCCCCTCCAGGCTTACCATCGGGTCATC 240

Qy        241 AGCTTGGAGGATTTCATGGAGAAGCTGGCACCCACCCACTGGCCCCCTGAGAAGCGGGTG 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db        241 AGCTTGGAGGATTTCATGGAGAAGCTGGCACCCACCCACTGGCCCCCTGAGAAGCGGGTG 300
Qy        301 GCATACTGCTTTGAGGTGGCAGCCCAGCGAAGCCCAGATAAGAAGACGTGCCCCATGAAG 360
Db        301 GCATACTGCTTTGAGGTGGCAGCCCAGCGAAGCCCAGATAAGAAGACGTGCCCCATGAAG 360
Qy        361 GAAGGAAACCCCTTTGGCCCATTCTGGGATCAGTTTCATGTGAGTTTCAACAAGTCGGAG 420
Db        361 GAAGGAAACCCCTTTGGCCCATTCTGGGATCAGTTTCATGTGAGTTTCAACAAGTCGGAG 420
Qy        421 CTTTTTACAGGCATTTCCTTCAGTGCTTCCTACAGAGAACAATGGAGCCAGAGATTTTCT 480
Db        421 CTTTTTACAGGCATTTCCTTCAGTGCTTCCTACAGAGAACAATGGAGCCAGAGATTTTCT 480
Qy        481 CCAAAGGAACATCCGGTGCTTGCCCTGCCAGGAGCCCCAGCCCAGTTCCCCGTCCTAGAA 540
Db        481 CCAAAGGAACATCCGGTGCTTGCCCTGCCAGGAGCCCCAGCCCAGTTCCCCGTCCTAGAA 540
Qy        541 GAACACAGGCCACTACAGAAGTACATGGTATGGTCAGACGAAATGGTGAAGACGGGAGAG 600
Db        541 GAACACAGGCCACTACAGAAGTACATGGTATGGTCAGACGAAATGGTGAAGACGGGAGAG 600
Qy        601 GCCCAGATTCATGCCCACCTTGTCCGGCCCTATGTGGGCATTCATCTGCGCATTGGCTCT 660
Db        601 GCCCAGATTCATGCCCACCTTGTCCGGCCCTATGTGGGCATTCATCTGCGCATTGGCTCT 660
Qy        661 GACTGGAAGAACGCCTGTGCCATGCTGAAGGACGGGACTGCAGGCTCGCACTTCATGGCC 720
Db        661 GACTGGAAGAACGCCTGTGCCATGCTGAAGGACGGGACTGCAGGCTCGCACTTCATGGCC 720
Qy        721 TCTCCGCAGTGTGTGGGCTACAGCCGCAGCACAGCGGCCCCCCTCACGATGACTATGTGC 780
Db        721 TCTCCGCAGTGTGTGGGCTACAGCCGCAGCACAGCGGCCCCCCTCACGATGACTATGTGC 780
Qy        781 CTGCCTGACCTGAAGGAGATCCAGAGGGCTGTGAAGCTCTGGGTGAGGTCGCTGGATGCC 840
Db        781 CTGCCTGACCTGAAGGAGATCCAGAGGGCTGTGAAGCTCTGGGTGAGGTCGCTGGATGCC 840
Qy        841 CAGTCGGTCTACGTTGCTACTGATTCCGAGAGTTATGTGCCTGAGCTCCAACAGCTCTTC 900
Db        841 CAGTCGGTCTACGTTGCTACTGATTCCGAGAGTTATGTGCCTGAGCTCCAACAGCTCTTC 900
Qy        901 AAAGGGAAGGTGAAGGTGGTGAGCCTGAAGCCTGAGGTGGCCCAGGTCGACCTGTACATC 960
Db        901 AAAGGGAAGGTGAAGGTGGTGAGCCTGAAGCCTGAGGTGGCCCAGGTCGACCTGTACATC 960
Qy        961 CTCGGCCAAGCCGACCACTTTATTGGCAACTGTGTCTCCTCCTTCACTGCCTTTGTGAAG 1020
Db        961 CTCGGCCAAGCCGACCACTTTATTGGCAACTGTGTCTCCTCCTTCACTGCCTTTGTGAAG 1020
Qy       1021 CGGGAGCGGGACCTCCAGGGGAGGCCGTCTTCTTTCTTCGGCATGGACAGGCCCCCTAAG 1080
Db       1021 CGGGAGCGGGACCTCCAGGGGAGGCCGTCTTCTTTCTTCGGCATGGACAGGCCCCCTAAG 1080
Qy       1081 CTGCGGGACGAGTTCTGATTCTGGCCGGAGCACCAGACCCTCTGATCCTGGAGGGACCAG 1140
Db       1081 CTGCGGGACGAGTTCTGATTCTGGCCGGAGCACCAGACCCTCTGATCCTGGAGGGACCAG 1140
Qy       1141 AGTCTGAGCTGGTCCTTCCAGCCAGGCCTGGCAGCCAGAGGTGCTCCGGGATTGCAAACT 1200
Db       1141 AGTCTGAGCTGGTCCTTCCAGCCAGGCCTGGCAGCCAGAGGTGCTCCGGGATTGCAAACT 1200
Qy       1201 CCTCTTCTCACCTGCCAAAGATGGAGAAGAGTGCCAGGGACCCCTCAAGGAGGGAGACGC 1260
Db       1201 CCTCTTCTCACCTGCCAAAGATGGAGAAGAGTGCCAGGGACCCCTCAAGGAGGGAGACGC 1260
Qy       1261 TCCATATCCCAGGGCATAGGACTTGCAGGTTCCTAGGAGCAGGAGCATCTCCCATCGCAC 1320
Db       1261 TCCATATCCCAGGGCATAGGACTTGCAGGTTCCTAGGAGCAGGAGCATCTCCCATCGCAC 1320
Qy       1321 GTGCTTTCTGCTCTTCTGGGAATTTCTCACACTGGCAAAGCAGTCCAGCCTCCGTCTTCT 1380
Db       1321 GTGCTTTCTGCTCTTCTGGGAATTTCTCACACTGGCAAAGCAGTCCAGCCTCCGTCTTCT 1380
```

```
Qy       1381 GGTCCACTCTGCTCTGAGCAGCCTGGGATGCTGAACTCTTCAGAGAGATTTTTTTATAGA 1440
Db       1381 GGTCCACTCTGCTCTGAGCAGCCTGGGATGCTGAACTCTTCAGAGAGATTTTTTTATAGA 1440
Qy       1441 GAGATTTCTATAATTTTGATACAAGGTCATGACTATCCTAGAACTCTCTGTGGTTTTTGA 1500
Db       1441 GAGATTTCTATAATTTTGATACAAGGTCATGACTATCCTAGAACTCTCTGTGGTTTTTGA 1500
Qy       1501 AAATCATTGAATTC 1514
Db       1501 AAATCATTGAATTC 1514
```

RESULT 2
BD103188

```
LOCUS       BD103188                1514 bp    DNA     linear   PAT 27-AUG-2002
DEFINITION  O-fucosyltransferase.
ACCESSION   BD103188
VERSION     BD103188.1  GI:22648762
KEYWORDS    JP 2001527389-A/1.
SOURCE      unidentified
  ORGANISM  unidentified
            unclassified.
REFERENCE   1  (bases 1 to 1514)
  AUTHORS   Wang,Y. and Spellman,M.W.
  TITLE     O-fucosyltransferase
  JOURNAL   Patent: JP 2001527389-A 1 25-DEC-2001;
            GENENTECH INC
COMMENT     OS   Unidentified
            PN   JP 2001527389-A/1
            PD   25-DEC-2001
            PF   17-DEC-1997 JP 1998532877
            PR   31-JAN-1997 US    08/792498,26-NOV-1997 US    08/978741 PI
            YANG WANG,MICHAEL W SPELLMAN
            PC   C12N15/54,C12N9/10,C07K16/40
            CC   Strandedness: Single;
            CC   Topology: Linear;
            CC   O-fucosyltransferase
            FH   Key             Location/Qualifiers
            FT   source          1. .1514
            FT                   /organism='Unidentified'.
FEATURES             Location/Qualifiers
     source          1. .1514
                     /organism="unidentified"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:32644"
ORIGIN
```

Query Match 100.0%; Score 1514; DB 6; Length 1514;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1514; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 ATGCCCGCGGGCTCCTGGGACCCGGCCGGTTACCTGCTCTACTGCCCCTGCATGGGGCGC 60
Db          1 ATGCCCGCGGGCTCCTGGGACCCGGCCGGTTACCTGCTCTACTGCCCCTGCATGGGGCGC 60
Qy         61 TTTGGGAACCAGGCCGATCACTTCTTGGGCTCTCTGGCATTTGCAAAGCTGCTAAACCGT 120
Db         61 TTTGGGAACCAGGCCGATCACTTCTTGGGCTCTCTGGCATTTGCAAAGCTGCTAAACCGT 120
Qy        121 ACCTTGGCTGTCCCTCCTTGGATTGAGTACCAGCATCACAAGCCTCCTTTCACCAACCTC 180
Db        121 ACCTTGGCTGTCCCTCCTTGGATTGAGTACCAGCATCACAAGCCTCCTTTCACCAACCTC 180
Qy        181 CATGTGTCCTACCAGAAGTACTTCAAGCTGGAGCCCCTCCAGGCTTACCATCGGGTCATC 240
Db        181 CATGTGTCCTACCAGAAGTACTTCAAGCTGGAGCCCCTCCAGGCTTACCATCGGGTCATC 240
Qy        241 AGCTTGGAGGATTTCATGGAGAAGCTGGCACCCACCCACTGGCCCCCTGAGAAGCGGGTG 300
Db        241 AGCTTGGAGGATTTCATGGAGAAGCTGGCACCCACCCACTGGCCCCCTGAGAAGCGGGTG 300
```

```
Qy       301 GCATACTGCTTTGAGGTGGCAGCCCAGCGAAGCCCAGATAAGAAGACGTGCCCCATGAAG 360
Db       301 GCATACTGCTTTGAGGTGGCAGCCCAGCGAAGCCCAGATAAGAAGACGTGCCCCATGAAG 360

Qy       361 GAAGGAAACCCCTTTGGCCCATTCTGGGATCAGTTTCATGTGAGTTTCAACAAGTCGGAG 420
Db       361 GAAGGAAACCCCTTTGGCCCATTCTGGGATCAGTTTCATGTGAGTTTCAACAAGTCGGAG 420

Qy       421 CTTTTTACAGGCATTTCCTTCAGTGCTTCCTACAGAGAACAATGGAGCCAGAGATTTTCT 480
Db       421 CTTTTTACAGGCATTTCCTTCAGTGCTTCCTACAGAGAACAATGGAGCCAGAGATTTTCT 480

Qy       481 CCAAAGGAACATCCGGTGCTTGCCCTGCCAGGAGCCCCAGCCCAGTTCCCCGTCCTAGAA 540
Db       481 CCAAAGGAACATCCGGTGCTTGCCCTGCCAGGAGCCCCAGCCCAGTTCCCCGTCCTAGAA 540

Qy       541 GAACACAGGCCACTACAGAAGTACATGGTATGGTCAGACGAAATGGTGAAGACGGGAGAG 600
Db       541 GAACACAGGCCACTACAGAAGTACATGGTATGGTCAGACGAAATGGTGAAGACGGGAGAG 600

Qy       601 GCCCAGATTCATGCCCACCTTGTCCGGCCCTATGTGGGCATTCATCTGCGCATTGGCTCT 660
Db       601 GCCCAGATTCATGCCCACCTTGTCCGGCCCTATGTGGGCATTCATCTGCGCATTGGCTCT 660

Qy       661 GACTGGAAGAACGCCTGTGCCATGCTGAAGGACGGGACTGCAGGCTCGCACTTCATGGCC 720
Db       661 GACTGGAAGAACGCCTGTGCCATGCTGAAGGACGGGACTGCAGGCTCGCACTTCATGGCC 720

Qy       721 TCTCCGCAGTGTGTGGGCTACAGCCGCAGCACAGCGGCCCCCCTCACGATGACTATGTGC 780
Db       721 TCTCCGCAGTGTGTGGGCTACAGCCGCAGCACAGCGGCCCCCCTCACGATGACTATGTGC 780

Qy       781 CTGCCTGACCTGAAGGAGATCCAGAGGGCTGTGAAGCTCTGGGTGAGGTCGCTGGATGCC 840
Db       781 CTGCCTGACCTGAAGGAGATCCAGAGGGCTGTGAAGCTCTGGGTGAGGTCGCTGGATGCC 840

Qy       841 CAGTCGGTCTACGTTGCTACTGATTCCGAGAGTTATGTGCCTGAGCTCCAACAGCTCTTC 900
Db       841 CAGTCGGTCTACGTTGCTACTGATTCCGAGAGTTATGTGCCTGAGCTCCAACAGCTCTTC 900

Qy       901 AAAGGGAAGGTGAAGGTGGTGAGCCTGAAGCCTGAGGTGGCCCAGGTCGACCTGTACATC 960
Db       901 AAAGGGAAGGTGAAGGTGGTGAGCCTGAAGCCTGAGGTGGCCCAGGTCGACCTGTACATC 960

Qy       961 CTCGGCCAAGCCGACCACTTTATTGGCAACTGTGTCTCCTCCTTCACTGCCTTTGTGAAG 1020
Db       961 CTCGGCCAAGCCGACCACTTTATTGGCAACTGTGTCTCCTCCTTCACTGCCTTTGTGAAG 1020

Qy      1021 CGGGAGCGGGACCTCCAGGGGAGGCCGTCTTCTTTCTTCGGCATGGACAGGCCCCCTAAG 1080
Db      1021 CGGGAGCGGGACCTCCAGGGGAGGCCGTCTTCTTTCTTCGGCATGGACAGGCCCCCTAAG 1080

Qy      1081 CTGCGGGACGAGTTCTGATTCTGGCCGGAGCACCAGACCCTCTGATCCTGGAGGGACCAG 1140
Db      1081 CTGCGGGACGAGTTCTGATTCTGGCCGGAGCACCAGACCCTCTGATCCTGGAGGGACCAG 1140

Qy      1141 AGTCTGAGCTGGTCCTTCCAGCCAGGCCTGGCAGCCAGAGGTGCTCCGGGATTGCAAACT 1200
Db      1141 AGTCTGAGCTGGTCCTTCCAGCCAGGCCTGGCAGCCAGAGGTGCTCCGGGATTGCAAACT 1200

Qy      1201 CCTCTTCTCACCTGCCAAAGATGGAGAAGAGTGCCAGGGACCCCTCAAGGAGGGAGACGC 1260
Db      1201 CCTCTTCTCACCTGCCAAAGATGGAGAAGAGTGCCAGGGACCCCTCAAGGAGGGAGACGC 1260

Qy      1261 TCCATATCCCAGGGCATAGGACTTGCAGGTTCCTAGGAGCAGGAGCATCTCCCATCGCAC 1320
Db      1261 TCCATATCCCAGGGCATAGGACTTGCAGGTTCCTAGGAGCAGGAGCATCTCCCATCGCAC 1320

Qy      1321 GTGCTTTCTGCTCTTCTGGGAATTTCTCACACTGGCAAAGCAGTCCAGCCTCCGTCTTCT 1380
Db      1321 GTGCTTTCTGCTCTTCTGGGAATTTCTCACACTGGCAAAGCAGTCCAGCCTCCGTCTTCT 1380

Qy      1381 GGTCCACTCTGCTCTGAGCAGCCTGGGATGCTGAACTCTTCAGAGAGATTTTTTTATAGA 1440
Db      1381 GGTCCACTCTGCTCTGAGCAGCCTGGGATGCTGAACTCTTCAGAGAGATTTTTTTATAGA 1440

Qy      1441 GAGATTTCTATAATTTTGATACAAGGTCATGACTATCCTAGAACTCTCTGTGGTTTTTGA 1500
Db      1441 GAGATTTCTATAATTTTGATACAAGGTCATGACTATCCTAGAACTCTCTGTGGTTTTTGA 1500

Qy      1501 AAATCATTGAATTC 1514
Db      1501 AAATCATTGAATTC 1514
```

RESULT 3
AF375884
LOCUS AF375884 5249 bp mRNA linear PRI 23-OCT-2001
DEFINITION Homo sapiens protein o-fucosyltransferase (POFUT1) mRNA, complete cds.
ACCESSION AF375884
VERSION AF375884.1 GI:15825113
KEYWORDS .
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 5249)
AUTHORS Wang,Y., Shao,L., Shi,S., Harris,R.J., Spellman,M.W., Stanley,P. and Haltiwanger,R.S.
TITLE Modification of epidermal growth factor-like repeats with O-fucose. Molecular cloning and expression of a novel GDP-fucose protein O-fucosyltransferase
JOURNAL J. Biol. Chem. 276 (43), 40338-40345 (2001)
MEDLINE 21523965
PUBMED 11524432
REFERENCE 2 (bases 1 to 5249)
AUTHORS Shi,S., Stanley,P., Wang,Y., Shao,L., Harris,R.J., Spellman,M.W. and Haltiwanger,R.S.
TITLE Direct Submission
JOURNAL Submitted (02-MAY-2001) Cell Biology, Albert Einstein College of Medicine, 1300 Morris Park Ave., New York, NY 10461, USA
FEATURES Location/Qualifiers
source 1. .5249
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/chromosome="20"
/map="between PLAGL2 and KIF3B"
/tissue_type="heart"
gene 1. .5249
/gene="POFUT1"
CDS 50. .1216
/gene="POFUT1"
/function="catalyzes transfer of fucose from GDP-fucose to serine or threonine on epidermal growth factor-like repeats"
/note="O-FucT-1; glycosyltransferase"
/codon_start=1
/product="protein o-fucosyltransferase"
/protein_id="AAL09576.1"
/db_xref="GI:15825114"
/translation="MGAAAWARPLSVSFLLLLLPLPGMPAGSWDPAGYLLYCPCMGRF
GNQADHFLGSLAFAKLLNRTLAVPPWIEYQHHKPPFTNLHVSYQKYFKLEPLQAYHRV
ISLEDFMEKLAPTHWPPEKRVAYCFEVAAQRSPDKKTCPMKEGNPFGPFWDQFHVSFN
KSELFTGISFSASYREQWSQRFSPKEHPVLALPGAPAQFPVLEEHRPLQKYMVWSDEM
VKTGEAQIHAHLVRPYVGIHLRIGSDWKNACAMLKDGTAGSHFMASPQCVGYSRSTAA
PLTMTMCLPDLKEIQRAVKLWVRSLDAQSVYVATDSESYVPELQQLFKGKVKVVSLKP
EVAQVDLYILGQADHFIGNCVSSFTAFVKRERDLQGRPSSFFGMDRPPKLRDEF"
ORIGIN

```
Query Match 99.9%; Score 1512.4; DB 9; Length 5249;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 1513; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 ATGCCCGCGGGCTCCTGGGACCCGGCCGGTTACCTGCTCTACTGCCCCTGCATGGGGCGC 60
```

```
Db        119 ATGCCCGCGGGCTCCTGGGACCCGGCCGGTTACCTGCTCTACTGCCCCTGCATGGGGCGC 178
Qy         61 TTTGGGAACCAGGCCGATCACTTCTTGGGCTCTCTGGCATTTGCAAAGCTGCTAAACCGT 120
Db        179 TTTGGGAACCAGGCCGATCACTTCTTGGGCTCTCTGGCATTTGCAAAGCTGCTAAACCGT 238
Qy        121 ACCTTGGCTGTCCCTCCTTGGATTGAGTACCAGCATCACAAGCCTCCTTTCACCAACCTC 180
Db        239 ACCTTGGCTGTCCCTCCTTGGATTGAGTACCAGCATCACAAGCCTCCTTTCACCAACCTC 298
Qy        181 CATGTGTCCTACCAGAAGTACTTCAAGCTGGAGCCCCTCCAGGCTTACCATCGGGTCATC 240
Db        299 CATGTGTCCTACCAGAAGTACTTCAAGCTGGAGCCCCTCCAGGCTTACCATCGGGTCATC 358
Qy        241 AGCTTGGAGGATTTCATGGAGAAGCTGGCACCCACCCACTGGCCCCCTGAGAAGCGGGTG 300
Db        359 AGCTTGGAGGATTTCATGGAGAAGCTGGCACCCACCCACTGGCCCCCTGAGAAGCGGGTG 418
Qy        301 GCATACTGCTTTGAGGTGGCAGCCCAGCGAAGCCCAGATAAGAAGACGTGCCCCATGAAG 360
Db        419 GCATACTGCTTTGAGGTGGCAGCCCAGCGAAGCCCAGATAAGAAGACGTGCCCCATGAAG 478
Qy        361 GAAGGAAACCCCTTTGGCCCATTCTGGGATCAGTTTCATGTGAGTTTCAACAAGTCGGAG 420
Db        479 GAAGGAAACCCCTTTGGCCCATTCTGGGATCAGTTTCATGTGAGTTTCAACAAGTCGGAG 538
Qy        421 CTTTTTACAGGCATTTCCTTCAGTGCTTCCTACAGAGAACAATGGAGCCAGAGATTTTCT 480
Db        539 CTTTTTACAGGCATTTCCTTCAGTGCTTCCTACAGAGAACAATGGAGCCAGAGATTTTCT 598
Qy        481 CCAAAGGAACATCCGGTGCTTGCCCTGCCAGGAGCCCCAGCCCAGTTCCCCGTCCTAGAA 540
Db        599 CCAAAGGAACATCCGGTGCTTGCCCTGCCAGGAGCCCCAGCCCAGTTCCCCGTCCTAGAG 658
Qy        541 GAACACAGGCCACTACAGAAGTACATGGTATGGTCAGACGAAATGGTGAAGACGGGAGAG 600
Db        659 GAACACAGGCCACTACAGAAGTACATGGTATGGTCAGACGAAATGGTGAAGACGGGAGAG 718
Qy        601 GCCCAGATTCATGCCCACCTTGTCCGGCCCTATGTGGGCATTCATCTGCGCATTGGCTCT 660
Db        719 GCCCAGATTCATGCCCACCTTGTCCGGCCCTATGTGGGCATTCATCTGCGCATTGGCTCT 778
Qy        661 GACTGGAAGAACGCCTGTGCCATGCTGAAGGACGGGACTGCAGGCTCGCACTTCATGGCC 720
Db        779 GACTGGAAGAACGCCTGTGCCATGCTGAAGGACGGGACTGCAGGCTCGCACTTCATGGCC 838
Qy        721 TCTCCGCAGTGTGTGGGCTACAGCCGCAGCACAGCGGCCCCCCTCACGATGACTATGTGC 780
Db        839 TCTCCGCAGTGTGTGGGCTACAGCCGCAGCACAGCGGCCCCCCTCACGATGACTATGTGC 898
Qy        781 CTGCCTGACCTGAAGGAGATCCAGAGGGCTGTGAAGCTCTGGGTGAGGTCGCTGGATGCC 840
Db        899 CTGCCTGACCTGAAGGAGATCCAGAGGGCTGTGAAGCTCTGGGTGAGGTCGCTGGATGCC 958
Qy        841 CAGTCGGTCTACGTTGCTACTGATTCCGAGAGTTATGTGCCTGAGCTCCAACAGCTCTTC 900
Db        959 CAGTCGGTCTACGTTGCTACTGATTCCGAGAGTTATGTGCCTGAGCTCCAACAGCTCTTC 1018
Qy        901 AAAGGGAAGGTGAAGGTGGTGAGCCTGAAGCCTGAGGTGGCCCAGGTCGACCTGTACATC 960
Db       1019 AAAGGGAAGGTGAAGGTGGTGAGCCTGAAGCCTGAGGTGGCCCAGGTCGACCTGTACATC 1078
Qy        961 CTCGGCCAAGCCGACCACTTTATTGGCAACTGTGTCTCCTCCTTCACTGCCTTTGTGAAG 1020
Db       1079 CTCGGCCAAGCCGACCACTTTATTGGCAACTGTGTCTCCTCCTTCACTGCCTTTGTGAAG 1138
Qy       1021 CGGGAGCGGGACCTCCAGGGGAGGCCGTCTTCTTTCTTCGGCATGGACAGGCCCCCTAAG 1080
Db       1139 CGGGAGCGGGACCTCCAGGGGAGGCCGTCTTCTTTCTTCGGCATGGACAGGCCCCCTAAG 1198
Qy       1081 CTGCGGGACGAGTTCTGATTCTGGCCGGAGCACCAGACCCTCTGATCCTGGAGGGACCAG 1140
Db       1199 CTGCGGGACGAGTTCTGATTCTGGCCGGAGCACCAGACCCTCTGATCCTGGAGGGACCAG 1258
Qy       1141 AGTCTGAGCTGGTCCTTCCAGCCAGGCCTGGCAGCCAGAGGTGCTCCGGGATTGCAAACT 1200
Db       1259 AGTCTGAGCTGGTCCTTCCAGCCAGGCCTGGCAGCCAGAGGTGCTCCGGGATTGCAAACT 1318
Qy       1201 CCTCTTCTCACCTGCCAAAGATGGAGAAGAGTGCCAGGGACCCCTCAAGGAGGGAGACGC 1260
Db       1319 CCTCTTCTCACCTGCCAAAGATGGAGAAGAGTGCCAGGGACCCCTCAAGGAGGGAGACGC 1378
Qy       1261 TCCATATCCCAGGGCATAGGACTTGCAGGTTCCTAGGAGCAGGAGCATCTCCCATCGCAC 1320
Db       1379 TCCATATCCCAGGGCATAGGACTTGCAGGTTCCTAGGAGCAGGAGCATCTCCCATCGCAC 1438
Qy       1321 GTGCTTTCTGCTCTTCTGGGAATTTCTCACACTGGCAAAGCAGTCCAGCCTCCGTCTTCT 1380
Db       1439 GTGCTTTCTGCTCTTCTGGGAATTTCTCACACTGGCAAAGCAGTCCAGCCTCCGTCTTCT 1498
Qy       1381 GGTCCACTCTGCTCTGAGCAGCCTGGGATGCTGAACTCTTCAGAGAGATTTTTTTATAGA 1440
Db       1499 GGTCCACTCTGCTCTGAGCAGCCTGGGATGCTGAACTCTTCAGAGAGATTTTTTTATAGA 1558
Qy       1441 GAGATTTCTATAATTTTGATACAAGGTCATGACTATCCTAGAACTCTCTGTGGTTTTTGA 1500
Db       1559 GAGATTTCTATAATTTTGATACAAGGTCATGACTATCCTAGAACTCTCTGTGGTTTTTGA 1618
Qy       1501 AAATCATTGAATTC 1514
Db       1619 AAATCATTGAATTC 1632
```

```
RESULT 4
BD103190
LOCUS       BD103190               11284 bp    DNA     linear   PAT 27-AUG-2002
DEFINITION  O-fucosyltransferase.
ACCESSION   BD103190
VERSION     BD103190.1  GI:22648764
KEYWORDS    JP 2001527389-A/3.
SOURCE      unidentified
  ORGANISM  unidentified
            unclassified.
REFERENCE   1  (bases 1 to 11284)
  AUTHORS   Wang,Y. and Spellman,M.W.
  TITLE     O-fucosyltransferase
  JOURNAL   Patent: JP 2001527389-A 3 25-DEC-2001;
            GENENTECH INC
COMMENT     OS   Unidentified
            PN   JP 2001527389-A/3
            PD   25-DEC-2001
            PF   17-DEC-1997 JP 1998532877
            PR   31-JAN-1997 US    08/792498,26-NOV-1997 US    08/978741 PI
            YANG WANG,MICHAEL W SPELLMAN
            PC   C12N15/54,C12N9/10,C07K16/40
            CC   Strandedness: Single;
            CC   Topology: Linear;
            CC   O-fucosyltransferase
            FH   Key             Location/Qualifiers
            FT   source          1. .11284
            FT                   /organism='Unidentified'.
FEATURES             Location/Qualifiers
     source          1. .11284
                     /organism="unidentified"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:32644"
ORIGIN

  Query Match           99.9%;  Score 1512.4;  DB 6;  Length 11284;
  Best Local Similarity   99.9%;  Pred. No. 0;
  Matches 1513;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 ATGCCCGCGGGCTCCTGGGACCCGGCCGGTTACCTGCTCTACTGCCCCTGCATGGGGCGC 60
Db       4236 ATGCCCGCGGGCTCCTGGGACCCGGCCGGTTACCTGCTCTACTGCCCCTGCATGGGGCGC 4295
```

```
Qy        61 TTTGGGAACCAGGCCGATCACTTCTTGGGCTCTCTGGCATTTGCAAAGCTGCTAAACCGT 120
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      4296 TTTGGGAACCAGGCCGATCACTTCTTGGGCTCTCTGGCATTTGCAAAGCTGCTAAACCGT 4355

Qy       121 ACCTTGGCTGTCCCTCCTTGGATTGAGTACCAGCATCACAAGCCTCCTTTCACCAACCTC 180
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      4356 ACCTTGGCTGTCCCTCCTTGGATTGAGTACCAGCATCACAAGCCTCCTTTCACCAACCTC 4415

Qy       181 CATGTGTCCTACCAGAAGTACTTCAAGCTGGAGCCCCTCCAGGCTTACCATCGGGTCATC 240
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      4416 CATGTGTCCTACCAGAAGTACTTCAAGCTGGAGCCCCTCCAGGCTTACCATCGGGTCATC 4475

Qy       241 AGCTTGGAGGATTTCATGGAGAAGCTGGCACCCACCCACTGGCCCCCTGAGAAGCGGGTG 300
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      4476 AGCTTGGAGGATTTCATGGAGAAGCTGGCACCCACCCACTGGCCCCCTGAGAAGCGGGTG 4535

Qy       301 GCATACTGCTTTGAGGTGGCAGCCCAGCGAAGCCCAGATAAGAAGACGTGCCCCATGAAG 360
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      4536 GCATACTGCTTTGAGGTGGCAGCCCAGCGAAGCCCAGATAAGAAGACGTGCCCCATGAAG 4595

Qy       361 GAAGGAAACCCCTTTGGCCCATTCTGGGATCAGTTTCATGTGAGTTTCAACAAGTCGGAG 420
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      4596 GAAGGAAACCCCTTTGGCCCATTCTGGGATCAGTTTCATGTGAGTTTCAACAAGTCGGAG 4655

Qy       421 CTTTTTACAGGCATTTCCTTCAGTGCTTCCTACAGAGAACAATGGAGCCAGAGATTTTCT 480
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      4656 CTTTTTACAGGCATTTCCTTCAGTGCTTCCTACAGAGAACAATGGAGCCAGAGATTTTCT 4715

Qy       481 CCAAAGGAACATCCGGTGCTTGCCCTGCCAGGAGCCCCAGCCCAGTTCCCCGTCCTAGAA 540
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      4716 CCAAAGGAACATCCGGTGCTTGCCCTGCCAGGAGCCCCAGCCCAGTTCCCCGTCCTAGAG 4775

Qy       541 GAACACAGGCCACTACAGAAGTACATGGTATGGTCAGACGAAATGGTGAAGACGGGAGAG 600
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      4776 GAACACAGGCCACTACAGAAGTACATGGTATGGTCAGACGAAATGGTGAAGACGGGAGAG 4835

Qy       601 GCCCAGATTCATGCCCACCTTGTCCGGCCCTATGTGGGCATTCATCTGCGCATTGGCTCT 660
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      4836 GCCCAGATTCATGCCCACCTTGTCCGGCCCTATGTGGGCATTCATCTGCGCATTGGCTCT 4895

Qy       661 GACTGGAAGAACGCCTGTGCCATGCTGAAGGACGGGACTGCAGGCTCGCACTTCATGGCC 720
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      4896 GACTGGAAGAACGCCTGTGCCATGCTGAAGGACGGGACTGCAGGCTCGCACTTCATGGCC 4955

Qy       721 TCTCCGCAGTGTGTGGGCTACAGCCGCAGCACAGCGGCCCCCCTCACGATGACTATGTGC 780
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      4956 TCTCCGCAGTGTGTGGGCTACAGCCGCAGCACAGCGGCCCCCCTCACGATGACTATGTGC 5015

Qy       781 CTGCCTGACCTGAAGGAGATCCAGAGGGCTGTGAAGCTCTGGGTGAGGTCGCTGGATGCC 840
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      5016 CTGCCTGACCTGAAGGAGATCCAGAGGGCTGTGAAGCTCTGGGTGAGGTCGCTGGATGCC 5075

Qy       841 CAGTCGGTCTACGTTGCTACTGATTCCGAGAGTTATGTGCCTGAGCTCCAACAGCTCTTC 900
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      5076 CAGTCGGTCTACGTTGCTACTGATTCCGAGAGTTATGTGCCTGAGCTCCAACAGCTCTTC 5135

Qy       901 AAAGGGAAGGTGAAGGTGGTGAGCCTGAAGCCTGAGGTGGCCCAGGTCGACCTGTACATC 960
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      5136 AAAGGGAAGGTGAAGGTGGTGAGCCTGAAGCCTGAGGTGGCCCAGGTCGACCTGTACATC 5195

Qy       961 CTCGGCCAAGCCGACCACTTTATTGGCAACTGTGTCTCCTCCTTCACTGCCTTTGTGAAG 1020
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      5196 CTCGGCCAAGCCGACCACTTTATTGGCAACTGTGTCTCCTCCTTCACTGCCTTTGTGAAG 5255

Qy      1021 CGGGAGCGGGACCTCCAGGGGAGGCCGTCTTCTTTCTTCGGCATGGACAGGCCCCCTAAG 1080
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      5256 CGGGAGCGGGACCTCCAGGGGAGGCCGTCTTCTTTCTTCGGCATGGACAGGCCCCCTAAG 5315

Qy      1081 CTGCGGGACGAGTTCTGATTCTGGCCGGAGCACCAGACCCTCTGATCCTGGAGGGACCAG 1140
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      5316 CTGCGGGACGAGTTCTGATTCTGGCCGGAGCACCAGACCCTCTGATCCTGGAGGGACCAG 5375

Qy      1141 AGTCTGAGCTGGTCCTTCCAGCCAGGCCTGGCAGCCAGAGGTGCTCCGGGATTGCAAACT 1200
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      5376 AGTCTGAGCTGGTCCTTCCAGCCAGGCCTGGCAGCCAGAGGTGCTCCGGGATTGCAAACT 5435

Qy      1201 CCTCTTCTCACCTGCCAAAGATGGAGAAGAGTGCCAGGGACCCCTCAAGGAGGGAGACGC 1260
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      5436 CCTCTTCTCACCTGCCAAAGATGGAGAAGAGTGCCAGGGACCCCTCAAGGAGGGAGACGC 5495

Qy      1261 TCCATATCCCAGGGCATAGGACTTGCAGGTTCCTAGGAGCAGGAGCATCTCCCATCGCAC 1320
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      5496 TCCATATCCCAGGGCATAGGACTTGCAGGTTCCTAGGAGCAGGAGCATCTCCCATCGCAC 5555

Qy      1321 GTGCTTTCTGCTCTTCTGGGAATTTCTCACACTGGCAAAGCAGTCCAGCCTCCGTCTTCT 1380
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      5556 GTGCTTTCTGCTCTTCTGGGAATTTCTCACACTGGCAAAGCAGTCCAGCCTCCGTCTTCT 5615

Qy      1381 GGTCCACTCTGCTCTGAGCAGCCTGGGATGCTGAACTCTTCAGAGAGATTTTTTTATAGA 1440
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      5616 GGTCCACTCTGCTCTGAGCAGCCTGGGATGCTGAACTCTTCAGAGAGATTTTTTTATAGA 5675

Qy      1441 GAGATTTCTATAATTTTGATACAAGGTCATGACTATCCTAGAACTCTCTGTGGTTTTTGA 1500
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      5676 GAGATTTCTATAATTTTGATACAAGGTCATGACTATCCTAGAACTCTCTGTGGTTTTTGA 5735

Qy      1501 AAATCATTGAATTC 1514
             ||||||||||||||
Db      5736 AAATCATTGAATTC 5749
```

```
RESULT 5
D80002
LOCUS       D80002                  5189 bp    mRNA    linear   PRI 10-JAN-2003
DEFINITION  Homo sapiens KIAA0180 mRNA, complete cds.
ACCESSION   D80002
VERSION     D80002.2  GI:20521837
KEYWORDS    .
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1
  AUTHORS   Nagase,T., Seki,N., Ishikawa,K., Tanaka,A. and Nomura,N.
  TITLE     Prediction of the coding sequences of unidentified human genes. V.
            The coding sequences of 40 new genes (KIAA0161-KIAA0200) deduced by
            analysis of cDNA clones from human cell line KG-1
  JOURNAL   DNA Res. 3 (1), 17-24 (1996)
  MEDLINE   96281124
   PUBMED   8724849
REFERENCE   2
  AUTHORS   Chiang,P.W., Wang,S., Smithivas,P., Song,W.J., Ramamoorthy,S.,
            Hillman,J., Puett,S., Van Keuren,M.L., Crombez,E., Kumar,A.,
            Glover,T.W., Miller,D.E., Tsai,C.H., Blackburn,C.C., Chen,X.N.,
            Sun,Z., Cheng,J.F., Korenberg,J.R. and Kurnit,D.M.
  TITLE     Identification and analysis of the human and murine putative
            chromatin structure regulator SUPT6H and Supt6h
  JOURNAL   Genomics 34 (3), 328-333 (1996)
  MEDLINE   96374824
   PUBMED   8786132
REFERENCE   3  (bases 1 to 5189)
  AUTHORS   Ohara,O., Nagase,T., Kikuno,R. and Nomura,N.
  TITLE     Direct Submission
  JOURNAL   Submitted (12-DEC-1995) Osamu Ohara, Kazusa DNA Research Institute;
            1532-3, Yana, Kisarazu, Chiba 292-0812, Japan
            (E-mail:cdnainfo@kazusa.or.jp, Tel:+81-438-52-3913)
COMMENT     On May 9, 2002 this sequence version replaced gi:1136419.
FEATURES             Location/Qualifiers
     source          1. .5189
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="ha02567s1"
                     /sex="male"
                     /cell_line="KG-1"
```

```
                     /cell_type="myeloblast"
                     /tissue_type="brain"
                     /clone_lib="pBluescriptII SK plus"
                     /note="This sequence was obtained by subcloning of the DNA
                     fragments derived from two cDNA clones (1 - 1066 was
                     derived from pf00933 and 1067 - 5189 was derived from
                     ha02567)."
     gene            1. .5189
                     /gene="KIAA0180"
     CDS             <1. .1213
                     /gene="KIAA0180"
                     /citation=[2]
                     /codon_start=2
                     /protein_id="BAA11497.2"
                     /db_xref="GI:20521838"
                     /translation="PPRLCAAAGSGSRADMGAAAWARPLSVSFLLLLLPLPGMPAGSW
                     DPAGYLLYCPCMGRFGNQADHFLGSLAFAKLLNRTLAVPPWIEYQHHKPPFTNLHVSY
                     QKYFKLEPLQAYHRVISLEDFMEKLAPTHWPPEKRVAYCFEVAAQRSPDKKTCPMKEG
                     NPFGPFWDQFHVSFNKSELFTGISFSASYREQWSQRFSPKEHPVLALPGAPAQFPVLE
                     EHRPLQKYMVWSDEMVKTGEAQIHAHLVRPYVGIHLRIGSDWKNACAMLKDGTAGSHF
                     MASPQCVGYSRSTAAPLTMTMCLPDLKEIQRAVKLWVRSLDAQSVYVATDSESYVPEL
                     QQLFKGKVKVVSLKPEVAQVDLYILGQADHFIGNCVSSFTAFVKRERDLQGRPSSFFG
                     MDRPPKLRDEF"
ORIGIN

  Query Match             99.8%;  Score 1510.8;  DB 9;  Length 5189;
  Best Local Similarity   99.9%;  Pred. No. 0;
  Matches 1512;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy          1 ATGCCCGCGGGCTCCTGGGACCCGGCCGGTTACCTGCTCTACTGCCCCTGCATGGGGCGC 60
              ||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        116 ATGCCTGCGGGCTCCTGGGACCCGGCCGGTTACCTGCTCTACTGCCCCTGCATGGGGCGC 175

Qy         61 TTTGGGAACCAGGCCGATCACTTCTTGGGCTCTCTGGCATTTGCAAAGCTGCTAAACCGT 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        176 TTTGGGAACCAGGCCGATCACTTCTTGGGCTCTCTGGCATTTGCAAAGCTGCTAAACCGT 235

Qy        121 ACCTTGGCTGTCCCTCCTTGGATTGAGTACCAGCATCACAAGCCTCCTTTCACCAACCTC 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        236 ACCTTGGCTGTCCCTCCTTGGATTGAGTACCAGCATCACAAGCCTCCTTTCACCAACCTC 295

Qy        181 CATGTGTCCTACCAGAAGTACTTCAAGCTGGAGCCCCTCCAGGCTTACCATCGGGTCATC 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        296 CATGTGTCCTACCAGAAGTACTTCAAGCTGGAGCCCCTCCAGGCTTACCATCGGGTCATC 355

Qy        241 AGCTTGGAGGATTTCATGGAGAAGCTGGCACCCACCCACTGGCCCCCTGAGAAGCGGGTG 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        356 AGCTTGGAGGATTTCATGGAGAAGCTGGCACCCACCCACTGGCCCCCTGAGAAGCGGGTG 415

Qy        301 GCATACTGCTTTGAGGTGGCAGCCCAGCGAAGCCCAGATAAGAAGACGTGCCCCATGAAG 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        416 GCATACTGCTTTGAGGTGGCAGCCCAGCGAAGCCCAGATAAGAAGACGTGCCCCATGAAG 475

Qy        361 GAAGGAAACCCCTTTGGCCCATTCTGGGATCAGTTTCATGTGAGTTTCAACAAGTCGGAG 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        476 GAAGGAAACCCCTTTGGCCCATTCTGGGATCAGTTTCATGTGAGTTTCAACAAGTCGGAG 535

Qy        421 CTTTTTACAGGCATTTCCTTCAGTGCTTCCTACAGAGAACAATGGAGCCAGAGATTTTCT 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        536 CTTTTTACAGGCATTTCCTTCAGTGCTTCCTACAGAGAACAATGGAGCCAGAGATTTTCT 595

Qy        481 CCAAAGGAACATCCGGTGCTTGCCCTGCCAGGAGCCCCAGCCCAGTTCCCCGTCCTAGAA 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        596 CCAAAGGAACATCCGGTGCTTGCCCTGCCAGGAGCCCCAGCCCAGTTCCCCGTCCTAGAG 655

Qy        541 GAACACAGGCCACTACAGAAGTACATGGTATGGTCAGACGAAATGGTGAAGACGGGAGAG 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        656 GAACACAGGCCACTACAGAAGTACATGGTATGGTCAGACGAAATGGTGAAGACGGGAGAG 715

Qy        601 GCCCAGATTCATGCCCACCTTGTCCGGCCCTATGTGGGCATTCATCTGCGCATTGGCTCT 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        716 GCCCAGATTCATGCCCACCTTGTCCGGCCCTATGTGGGCATTCATCTGCGCATTGGCTCT 775

Qy        661 GACTGGAAGAACGCCTGTGCCATGCTGAAGGACGGGACTGCAGGCTCGCACTTCATGGCC 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        776 GACTGGAAGAACGCCTGTGCCATGCTGAAGGACGGGACTGCAGGCTCGCACTTCATGGCC 835

Qy        721 TCTCCGCAGTGTGTGGGCTACAGCCGCAGCACAGCGGCCCCCCTCACGATGACTATGTGC 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        836 TCTCCGCAGTGTGTGGGCTACAGCCGCAGCACAGCGGCCCCCCTCACGATGACTATGTGC 895

Qy        781 CTGCCTGACCTGAAGGAGATCCAGAGGGCTGTGAAGCTCTGGGTGAGGTCGCTGGATGCC 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        896 CTGCCTGACCTGAAGGAGATCCAGAGGGCTGTGAAGCTCTGGGTGAGGTCGCTGGATGCC 955

Qy        841 CAGTCGGTCTACGTTGCTACTGATTCCGAGAGTTATGTGCCTGAGCTCCAACAGCTCTTC 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        956 CAGTCGGTCTACGTTGCTACTGATTCCGAGAGTTATGTGCCTGAGCTCCAACAGCTCTTC 1015

Qy        901 AAAGGGAAGGTGAAGGTGGTGAGCCTGAAGCCTGAGGTGGCCCAGGTCGACCTGTACATC 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1016 AAAGGGAAGGTGAAGGTGGTGAGCCTGAAGCCTGAGGTGGCCCAGGTCGACCTGTACATC 1075

Qy        961 CTCGGCCAAGCCGACCACTTTATTGGCAACTGTGTCTCCTCCTTCACTGCCTTTGTGAAG 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1076 CTCGGCCAAGCCGACCACTTTATTGGCAACTGTGTCTCCTCCTTCACTGCCTTTGTGAAG 1135

Qy       1021 CGGGAGCGGGACCTCCAGGGGAGGCCGTCTTCTTTCTTCGGCATGGACAGGCCCCCTAAG 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1136 CGGGAGCGGGACCTCCAGGGGAGGCCGTCTTCTTTCTTCGGCATGGACAGGCCCCCTAAG 1195

Qy       1081 CTGCGGGACGAGTTCTGATTCTGGCCGGAGCACCAGACCCTCTGATCCTGGAGGGACCAG 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1196 CTGCGGGACGAGTTCTGATTCTGGCCGGAGCACCAGACCCTCTGATCCTGGAGGGACCAG 1255

Qy       1141 AGTCTGAGCTGGTCCTTCCAGCCAGGCCTGGCAGCCAGAGGTGCTCCGGGATTGCAAACT 1200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1256 AGTCTGAGCTGGTCCTTCCAGCCAGGCCTGGCAGCCAGAGGTGCTCCGGGATTGCAAACT 1315

Qy       1201 CCTCTTCTCACCTGCCAAAGATGGAGAAGAGTGCCAGGGACCCCTCAAGGAGGGAGACGC 1260
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1316 CCTCTTCTCACCTGCCAAAGATGGAGAAGAGTGCCAGGGACCCCTCAAGGAGGGAGACGC 1375

Qy       1261 TCCATATCCCAGGGCATAGGACTTGCAGGTTCCTAGGAGCAGGAGCATCTCCCATCGCAC 1320
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1376 TCCATATCCCAGGGCATAGGACTTGCAGGTTCCTAGGAGCAGGAGCATCTCCCATCGCAC 1435

Qy       1321 GTGCTTTCTGCTCTTCTGGGAATTTCTCACACTGGCAAAGCAGTCCAGCCTCCGTCTTCT 1380
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1436 GTGCTTTCTGCTCTTCTGGGAATTTCTCACACTGGCAAAGCAGTCCAGCCTCCGTCTTCT 1495

Qy       1381 GGTCCACTCTGCTCTGAGCAGCCTGGGATGCTGAACTCTTCAGAGAGATTTTTTTATAGA 1440
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1496 GGTCCACTCTGCTCTGAGCAGCCTGGGATGCTGAACTCTTCAGAGAGATTTTTTTATAGA 1555

Qy       1441 GAGATTTCTATAATTTTGATACAAGGTCATGACTATCCTAGAACTCTCTGTGGTTTTTGA 1500
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1556 GAGATTTCTATAATTTTGATACAAGGTCATGACTATCCTAGAACTCTCTGTGGTTTTTGA 1615

Qy       1501 AAATCATTGAATTC 1514
              ||||||||||||||
Db       1616 AAATCATTGAATTC 1629
```

```
RESULT 6
AX780153
LOCUS       AX780153                5266 bp    DNA     linear   PAT 14-JUL-2003
DEFINITION  Sequence 2310 from Patent WO03039443.
ACCESSION   AX780153
VERSION     AX780153.1  GI:32697147
KEYWORDS    .
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1
```

```
  AUTHORS   Haferlach,T., Schoch,C., Kern,W., Kohlmann,A., Schnittger,S.,
            Dugas,M., Eils,R., Brors,B. and Mergenthaler,S.
  TITLE     Novel genetic markers for leukemias
  JOURNAL   Patent: WO 03039443-A 2310 15-MAY-2003;
            Deutsches Krebsforschungszentrum (DE) ;
            Ludwig-Maximilian-Universitaet Muenchen (DE) ; Haferlach, Torsten,
            PD Dr. Dr. (DE) ; Schoch, Claudia (DE) ; Kern, Wolfgang (DE)
FEATURES             Location/Qualifiers
     source          1. .5266
                     /organism="Homo sapiens"
                     /mol_type="unassigned DNA"
                     /db_xref="taxon:9606"
ORIGIN

  Query Match             99.7%;  Score 1509.8;  DB 6;  Length 5266;
  Best Local Similarity   99.8%;  Pred. No. 0;
  Matches 1511;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy          1 ATGCCCGCGGGCTCCTGGGACCCGGCCGGTTACCTGCTCTACTGCCCCTGCATGGGGCGC 60
              ||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        132 ATGCCTGCGGGCTCCTGGGACCCGGCCGGTTACCTGCTCTACTGCCCCTGCATGGGGCGC 191

Qy         61 TTTGGGAACCAGGCCGATCACTTCTTGGGCTCTCTGGCATTTGCAAAGCTGCTAAACCGT 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        192 TTTGGGAACCAGGCCGATCACTTCTTGGGCTCTCTGGCATTTGCAAAGCTGCTAAACCGT 251

Qy        121 ACCTTGGCTGTCCCTCCTTGGATTGAGTACCAGCATCACAAGCCTCCTTTCACCAACCTC 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        252 ACCTTGGCTGTCCCTCCTTGGATTGAGTACCAGCATCACAAGCCTCCTTTCACCAACCTC 311

Qy        181 CATGTGTCCTACCAGAAGTACTTCAAGCTGGAGCCCCTCCAGGCTTACCATCGGGTCATC 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        312 CATGTGTCCTACCAGAAGTACTTCAAGCTGGAGCCCCTCCAGGCTTACCATCGGGTCATC 371

Qy        241 AGCTTGGAGGATTTCATGGAGAAGCTGGCACCCACCCACTGGCCCCCTGAGAAGCGGGTG 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        372 AGCTTGGAGGATTTCATGGAGAAGCTGGCACCCACCCACTGGCCCCCTGAGAAGCGGGTG 431

Qy        301 GCATACTGCTTTGAGGTGGCAGCCCAGCGAAGCCCAGATAAGAAGACGTGCCCCATGAAG 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        432 GCATACTGCTTTGAGGTGGCAGCCCAGCGAAGCCCAGATAAGAAGACGTGCCCCATGAAG 491

Qy        361 GAAGGAAACCCCTTTGGCCCATTCTGGGATCAGTTTCATGTGAGTTTCAACAAGTCGGAG 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        492 GAAGGAAACCCCTTTGGCCCATTCTGGGATCAGTTTCATGTGAGTTTCAACAAGTCGGAG 551

Qy        421 CTTTTTACAGGCATTTCCTTCAGTGCTTCCTACAGAGAACAATGGAGCCAGAGATTTTCT 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        552 CTTTTTACAGGCATTTCCTTCAGTGCTTCCTACAGAGAACAATGGAGCCAGAGATTTTCT 611

Qy        481 CCAAAGGAACATCCGGTGCTTGCCCTGCCAGGAGCCCCAGCCCAGTTCCCCGTCCTAGAA 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        612 CCAAAGGAACATCCGGTGCTTGCCCTGCCAGGAGCCCCAGCCCAGTTCCCCGTCCTAGAG 671

Qy        541 GAACACAGGCCACTACAGAAGTACATGGTATGGTCAGACGAAATGGTGAAGACGGGAGAG 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        672 GAACACAGGCCACTACAGAAGTACATGGTATGGTCAGACGAAATGGTGAAGACGGGAGAG 731

Qy        601 GCCCAGATTCATGCCCACCTTGTCCGGCCCTATGTGGGCATTCATCTGCGCATTGGCTCT 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        732 GCCCAGATTCATGCCCACCTTGTCCGGCCCTATGTGGGCATTCATCTGCGCATTGGCTCT 791

Qy        661 GACTGGAAGAACGCCTGTGCCATGCTGAAGGACGGGACTGCAGGCTCGCACTTCATGGCC 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        792 GACTGGAAGAACGCCTGTGCCATGCTGAAGGACGGGACTGCAGGCTCGCACTTCATGGCC 851

Qy        721 TCTCCGCAGTGTGTGGGCTACAGCCGCAGCACAGCGGCCCCCCTCACGATGACTATGTGC 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        852 TCTCCGCAGTGTGTGGGCTACAGCCGCAGCACAGCGGCCCCCCTCACGATGACTATGTGC 911

Qy        781 CTGCCTGACCTGAAGGAGATCCAGAGGGCTGTGAAGCTCTGGGTGAGGTCGCTGGATGCC 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        912 CTGCCTGACCTGAAGGAGATCCAGAGGGCTGTGAAGCTCTGGGTGAGGTCGCTGGATGCC 971

Qy        841 CAGTCGGTCTACGTTGCTACTGATTCCGAGAGTTATGTGCCTGAGCTCCAACAGCTCTTC 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        972 CAGTCGGTCTACGTTGCTACTGATTCCGAGAGTTATGTGCCTGAGCTCCAACAGCTCTTC 1031

Qy        901 AAAGGGAAGGTGAAGGTGGTGAGCCTGAAGCCTGAGGTGGCCCAGGTCGACCTGTACATC 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1032 AAAGGGAAGGTGAAGGTGGTGAGCCTGAAGCCTGAGGTGGCCCAGGTCGACCTGTACATC 1091

Qy        961 CTCGGCCAAGCCGACCACTTTATTGGCAACTGTGTCTCCTCCTTCACTGCCTTTGTGAAG 1020
              ||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1092 CTCGGNCAAGCCGACCACTTTATTGGCAACTGTGTCTCCTCCTTCACTGCCTTTGTGAAG 1151

Qy       1021 CGGGAGCGGGACCTCCAGGGGAGGCCGTCTTCTTTCTTCGGCATGGACAGGCCCCCTAAG 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1152 CGGGAGCGGGACCTCCAGGGGAGGCCGTCTTCTTTCTTCGGCATGGACAGGCCCCCTAAG 1211

Qy       1081 CTGCGGGACGAGTTCTGATTCTGGCCGGAGCACCAGACCCTCTGATCCTGGAGGGACCAG 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1212 CTGCGGGACGAGTTCTGATTCTGGCCGGAGCACCAGACCCTCTGATCCTGGAGGGACCAG 1271

Qy       1141 AGTCTGAGCTGGTCCTTCCAGCCAGGCCTGGCAGCCAGAGGTGCTCCGGGATTGCAAACT 1200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1272 AGTCTGAGCTGGTCCTTCCAGCCAGGCCTGGCAGCCAGAGGTGCTCCGGGATTGCAAACT 1331

Qy       1201 CCTCTTCTCACCTGCCAAAGATGGAGAAGAGTGCCAGGGACCCCTCAAGGAGGGAGACGC 1260
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1332 CCTCTTCTCACCTGCCAAAGATGGAGAAGAGTGCCAGGGACCCCTCAAGGAGGGAGACGC 1391

Qy       1261 TCCATATCCCAGGGCATAGGACTTGCAGGTTCCTAGGAGCAGGAGCATCTCCCATCGCAC 1320
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1392 TCCATATCCCAGGGCATAGGACTTGCAGGTTCCTAGGAGCAGGAGCATCTCCCATCGCAC 1451

Qy       1321 GTGCTTTCTGCTCTTCTGGGAATTTCTCACACTGGCAAAGCAGTCCAGCCTCCGTCTTCT 1380
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1452 GTGCTTTCTGCTCTTCTGGGAATTTCTCACACTGGCAAAGCAGTCCAGCCTCCGTCTTCT 1511

Qy       1381 GGTCCACTCTGCTCTGAGCAGCCTGGGATGCTGAACTCTTCAGAGAGATTTTTTTATAGA 1440
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1512 GGTCCACTCTGCTCTGAGCAGCCTGGGATGCTGAACTCTTCAGAGAGATTTTTTTATAGA 1571

Qy       1441 GAGATTTCTATAATTTTGATACAAGGTCATGACTATCCTAGAACTCTCTGTGGTTTTTGA 1500
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1572 GAGATTTCTATAATTTTGATACAAGGTCATGACTATCCTAGAACTCTCTGTGGTTTTTGA 1631

Qy       1501 AAATCATTGAATTC 1514
              ||||||||||||||
Db       1632 AAATCATTGAATTC 1645


RESULT 7
HSM803330
LOCUS       HSM803330              4560 bp    mRNA    linear   PRI 13-MAY-2003
DEFINITION  Homo sapiens mRNA; cDNA DKFZp451J114 (from clone DKFZp451J114).
ACCESSION   AL832023
VERSION     AL832023.1  GI:21732563
KEYWORDS    .
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 4560)
  AUTHORS   Wambutt,R., Heubner,D., Mewes,H.W., Weil,B., Amid,C., Osanger,A.,
            Fobo,G., Han,M. and Wiemann,S.
  TITLE     Direct Submission
  JOURNAL   Submitted (13-MAY-2003) MIPS, Ingolstaedter Landstr.1, D-85764
            Neuherberg, GERMANY
COMMENT     Clone from S. Wiemann, Molecular Genome Analysis, German Cancer
            Research Center (DKFZ); Email s.wiemann@dkfz-heidelberg.de;
            sequenced by AGOWA (Berlin/Germany) within the cDNA sequencing
            consortium of the German Genome Project.
            This clone (DKFZp451J114) is available at the RZPD in Berlin.
            Please contact the RZPD: Ressourcenzentrum, Heubnerweg 6, 14059
```

```
           Berlin-Charlottenburg, GERMANY; Email: clone@rzpd.de Further
           information about the clone and the sequencing project is available
           at http://mips.gsf.de/proj/cDNA/.
FEATURES             Location/Qualifiers
     source          1. .4560
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="RZPD:DKFZp451J114"
                     /db_xref="taxon:9606"
                     /clone="DKFZp451J114"
                     /tissue_type="human skeletal muscle"
                     /clone_lib="451 (synonym: hlcc1). Vector pSport1; host
                     DH10B; sites NotI + SalI"
                     /dev_stage="adult"
     polyA_site      4538
ORIGIN

  Query Match             99.6%;  Score 1507.6;  DB 9;  Length 4560;
  Best Local Similarity   99.7%;  Pred. No. 0;
  Matches 1510;  Conservative    0;  Mismatches    4;  Indels    0;  Gaps    0;

Qy          1 ATGCCCGCGGGCTCCTGGGACCCGGCCGGTTACCTGCTCTACTGCCCCTGCATGGGGCGC 60
              ||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         51 ATGCCTGCGGGCTCCTGGGACCCGGCCGGTTACCTGCTCTACTGCCCCTGCATGGGGCGC 110

Qy         61 TTTGGGAACCAGGCCGATCACTTCTTGGGCTCTCTGGCATTTGCAAAGCTGCTAAACCGT 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        111 TTTGGGAACCAGGCCGATCACTTCTTGGGCTCTCTGGCATTTGCAAAGCTGCTAAACCGT 170

Qy        121 ACCTTGGCTGTCCCTCCTTGGATTGAGTACCAGCATCACAAGCCTCCTTTCACCAACCTC 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        171 ACCTTGGCTGTCCCTCCTTGGATTGAGTACCAGCATCACAAGCCTCCTTTCACCAACCTC 230

Qy        181 CATGTGTCCTACCAGAAGTACTTCAAGCTGGAGCCCCTCCAGGCTTACCATCGGGTCATC 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        231 CATGTGTCCTACCAGAAGTACTTCAAGCTGGAGCCCCTCCAGGCTTACCATCGGGTCATC 290

Qy        241 AGCTTGGAGGATTTCATGGAGAAGCTGGCACCCACCCACTGGCCCCCTGAGAAGCGGGTG 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        291 AGCTTGGAGGATTTCATGGAGAAGCTGGCACCCACCCACTGGCCCCCTGAGAAGCGGGTG 350

Qy        301 GCATACTGCTTTGAGGTGGCAGCCCAGCGAAGCCCAGATAAGAAGACGTGCCCCATGAAG 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        351 GCATACTGCTTTGAGGTGGCAGCCCAGCGAAGCCCAGATAAGAAGACGTGCCCCATGAAG 410

Qy        361 GAAGGAAACCCCTTTGGCCCATTCTGGGATCAGTTTCATGTGAGTTTCAACAAGTCGGAG 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        411 GAAGGAAACCCCTTTGGCCCATTCTGGGATCAGTTTCATGTGAGTTTCAACAAGTCGGAG 470

Qy        421 CTTTTTACAGGCATTTCCTTCAGTGCTTCCTACAGAGAACAATGGAGCCAGAGATTTTCT 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        471 CTTTTTACAGGCATTTCCTTCAGTGCTTCCTACAGAGAACAATGGAGCCAGAGATTTTCT 530

Qy        481 CCAAAGGAACATCCGGTGCTTGCCCTGCCAGGAGCCCCAGCCCAGTTCCCCGTCCTAGAA 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        531 CCAAAGGAACATCCGGTGCTTGCCCTGCCAGGAGCCCCAGCCCAGTTCCCCGTCCTAGAG 590

Qy        541 GAACACAGGCCACTACAGAAGTACATGGTATGGTCAGACGAAATGGTGAAGACGGGAGAG 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        591 GAACACAGGCCACTACAGAAGTACATGGTATGGTCAGACGAAATGGTGAAGACGGGAGAG 650

Qy        601 GCCCAGATTCATGCCCACCTTGTCCGGCCCTATGTGGGCATTCATCTGCGCATTGGCTCT 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        651 GCCCAGATTCATGCCCACCTTGTCCGGCCCTATGTGGGCATTCATCTGCGCATTGGCTCT 710

Qy        661 GACTGGAAGAACGCCTGTGCCATGCTGAAGGACGGGACTGCAGGCTCGCACTTCATGGCC 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        711 GACTGGAAGAACGCCTGTGCCATGCTGAAGGACGGGACTGCAGGCTCGCACTTCATGGCC 770

Qy        721 TCTCCGCAGTGTGTGGGCTACAGCCGCAGCACAGCGGCCCCCCTCACGATGACTATGTGC 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        771 TCTCCGCAGTGTGTGGGCTACAGCCGCAGCACAGCGGCCCCCCTCACGATGACTATGTGC 830

Qy        781 CTGCCTGACCTGAAGGAGATCCAGAGGGCTGTGAAGCTCTGGGTGAGGTCGCTGGATGCC 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        831 CTGCCTGACCTGAAGGAGATCCAGAGGGCTGTGAAGCTCTGGGTGAGGTCGCTGGATGCC 890

Qy        841 CAGTCGGTCTACGTTGCTACTGATTCCGAGAGTTATGTGCCTGAGCTCCAACAGCTCTTC 900
              |||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||
Db        891 CAGTCGGTCTACGTTGCTACCGATTCCGAGAGTTATGTGCCTGAGCTCCAACAGCTCTTC 950

Qy        901 AAAGGGAAGGTGAAGGTGGTGAGCCTGAAGCCTGAGGTGGCCCAGGTCGACCTGTACATC 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        951 AAAGGGAAGGTGAAGGTGGTGAGCCTGAAGCCTGAGGTGGCCCAGGTCGACCTGTACATC 1010

Qy        961 CTCGGCCAAGCCGACCACTTTATTGGCAACTGTGTCTCCTCCTTCACTGCCTTTGTGAAG 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1011 CTCGGCCAAGCCGACCACTTTATTGGCAACTGTGTCTCCTCCTTCACTGCCTTTGTGAAG 1070

Qy       1021 CGGGAGCGGGACCTCCAGGGGAGGCCGTCTTCTTTCTTCGGCATGGACAGGCCCCCTAAG 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1071 CGGGAGCGGGACCTCCAGGGGAGGCCGTCTTCTTTCTTCGGCATGGACAGGCCCCCTAAG 1130

Qy       1081 CTGCGGGACGAGTTCTGATTCTGGCCGGAGCACCAGACCCTCTGATCCTGGAGGGACCAG 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1131 CTGCGGGACGAGTTCTGATTCTGGCCGGAGCACCAGACCCTCTGATCCTGGAGGGACCAG 1190

Qy       1141 AGTCTGAGCTGGTCCTTCCAGCCAGGCCTGGCAGCCAGAGGTGCTCCGGGATTGCAAACT 1200
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||| ||
Db       1191 AGTCTGAGCTGGTCCTTCCAGCCAGGCCTGGCAGCCAGAGGTGCTCCGGGATTGCAAGCT 1250

Qy       1201 CCTCTTCTCACCTGCCAAAGATGGAGAAGAGTGCCAGGGACCCCTCAAGGAGGGAGACGC 1260
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1251 CCTCTTCTCACCTGCCAAAGATGGAGAAGAGTGCCAGGGACCCCTCAAGGAGGGAGACGC 1310

Qy       1261 TCCATATCCCAGGGCATAGGACTTGCAGGTTCCTAGGAGCAGGAGCATCTCCCATCGCAC 1320
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1311 TCCATATCCCAGGGCATAGGACTTGCAGGTTCCTAGGAGCAGGAGCATCTCCCATCGCAC 1370

Qy       1321 GTGCTTTCTGCTCTTCTGGGAATTTCTCACACTGGCAAAGCAGTCCAGCCTCCGTCTTCT 1380
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1371 GTGCTTTCTGCTCTTCTGGGAATTTCTCACACTGGCAAAGCAGTCCAGCCTCCGTCTTCT 1430

Qy       1381 GGTCCACTCTGCTCTGAGCAGCCTGGGATGCTGAACTCTTCAGAGAGATTTTTTTATAGA 1440
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1431 GGTCCACTCTGCTCTGAGCAGCCTGGGATGCTGAACTCTTCAGAGAGATTTTTTTATAGA 1490

Qy       1441 GAGATTTCTATAATTTTGATACAAGGTCATGACTATCCTAGAACTCTCTGTGGTTTTTGA 1500
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1491 GAGATTTCTATAATTTTGATACAAGGTCATGACTATCCTAGAACTCTCTGTGGTTTTTGA 1550

Qy       1501 AAATCATTGAATTC 1514
              ||||||||||||||
Db       1551 AAATCATTGAATTC 1564


RESULT 8
AR163453
LOCUS       AR163453                5009 bp    DNA     linear   PAT 17-OCT-2001
DEFINITION  Sequence 8 from patent US 6270987.
ACCESSION   AR163453
VERSION     AR163453.1  GI:16234058
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 5009)
  AUTHORS   Wang,Y. and Spellman,M.W.
  TITLE     O-fucosyltransferase
  JOURNAL   Patent: US 6270987-A 8 07-AUG-2001;
FEATURES             Location/Qualifiers
     source          1. .5009
                     /organism="unknown"
                     /mol_type="unassigned DNA"
ORIGIN
```

```
  Query Match             95.6%;  Score 1447.4;  DB 6;  Length 5009;
  Best Local Similarity   99.9%;  Pred. No. 0;
  Matches 1448;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy         66 GAACCAGGCCGATCACTTCTTGGGCTCTCTGGCATTTGCAAAGCTGCTAAACCGTACCTT 125
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 GAACCAGGCCGATCACTTCTTGGGCTCTCTGGCATTTGCAAAGCTGCTAAACCGTACCTT 60

Qy        126 GGCTGTCCCTCCTTGGATTGAGTACCAGCATCACAAGCCTCCTTTCACCAACCTCCATGT 185
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GGCTGTCCCTCCTTGGATTGAGTACCAGCATCACAAGCCTCCTTTCACCAACCTCCATGT 120

Qy        186 GTCCTACCAGAAGTACTTCAAGCTGGAGCCCCTCCAGGCTTACCATCGGGTCATCAGCTT 245
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 GTCCTACCAGAAGTACTTCAAGCTGGAGCCCCTCCAGGCTTACCATCGGGTCATCAGCTT 180

Qy        246 GGAGGATTTCATGGAGAAGCTGGCACCCACCCACTGGCCCCCTGAGAAGCGGGTGGCATA 305
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 GGAGGATTTCATGGAGAAGCTGGCACCCACCCACTGGCCCCCTGAGAAGCGGGTGGCATA 240

Qy        306 CTGCTTTGAGGTGGCAGCCCAGCGAAGCCCAGATAAGAAGACGTGCCCCATGAAGGAAGG 365
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 CTGCTTTGAGGTGGCAGCCCAGCGAAGCCCAGATAAGAAGACGTGCCCCATGAAGGAAGG 300

Qy        366 AAACCCCTTTGGCCCATTCTGGGATCAGTTTCATGTGAGTTTCAACAAGTCGGAGCTTTT 425
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 AAACCCCTTTGGCCCATTCTGGGATCAGTTTCATGTGAGTTTCAACAAGTCGGAGCTTTT 360

Qy        426 TACAGGCATTTCCTTCAGTGCTTCCTACAGAGAACAATGGAGCCAGAGATTTTCTCCAAA 485
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 TACAGGCATTTCCTTCAGTGCTTCCTACAGAGAACAATGGAGCCAGAGATTTTCTCCAAA 420

Qy        486 GGAACATCCGGTGCTTGCCCTGCCAGGAGCCCCAGCCCAGTTCCCCGTCCTAGAAGAACA 545
              |||||||||||||||||||||||||||||||||||||||||||||||||| |||||||||
Db        421 GGAACATCCGGTGCTTGCCCTGCCAGGAGCCCCAGCCCAGTTCCCCGTCCTAGAGGAACA 480

Qy        546 CAGGCCACTACAGAAGTACATGGTATGGTCAGACGAAATGGTGAAGACGGGAGAGGCCCA 605
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 CAGGCCACTACAGAAGTACATGGTATGGTCAGACGAAATGGTGAAGACGGGAGAGGCCCA 540

Qy        606 GATTCATGCCCACCTTGTCCGGCCCTATGTGGGCATTCATCTGCGCATTGGCTCTGACTG 665
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 GATTCATGCCCACCTTGTCCGGCCCTATGTGGGCATTCATCTGCGCATTGGCTCTGACTG 600

Qy        666 GAAGAACGCCTGTGCCATGCTGAAGGACGGGACTGCAGGCTCGCACTTCATGGCCTCTCC 725
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 GAAGAACGCCTGTGCCATGCTGAAGGACGGGACTGCAGGCTCGCACTTCATGGCCTCTCC 660

Qy        726 GCAGTGTGTGGGCTACAGCCGCAGCACAGCGGCCCCCCTCACGATGACTATGTGCCTGCC 785
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 GCAGTGTGTGGGCTACAGCCGCAGCACAGCGGCCCCCCTCACGATGACTATGTGCCTGCC 720

Qy        786 TGACCTGAAGGAGATCCAGAGGGCTGTGAAGCTCTGGGTGAGGTCGCTGGATGCCCAGTC 845
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 TGACCTGAAGGAGATCCAGAGGGCTGTGAAGCTCTGGGTGAGGTCGCTGGATGCCCAGTC 780

Qy        846 GGTCTACGTTGCTACTGATTCCGAGAGTTATGTGCCTGAGCTCCAACAGCTCTTCAAAGG 905
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 GGTCTACGTTGCTACTGATTCCGAGAGTTATGTGCCTGAGCTCCAACAGCTCTTCAAAGG 840

Qy        906 GAAGGTGAAGGTGGTGAGCCTGAAGCCTGAGGTGGCCCAGGTCGACCTGTACATCCTCGG 965
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 GAAGGTGAAGGTGGTGAGCCTGAAGCCTGAGGTGGCCCAGGTCGACCTGTACATCCTCGG 900

Qy        966 CCAAGCCGACCACTTTATTGGCAACTGTGTCTCCTCCTTCACTGCCTTTGTGAAGCGGGA 1025
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 CCAAGCCGACCACTTTATTGGCAACTGTGTCTCCTCCTTCACTGCCTTTGTGAAGCGGGA 960

Qy       1026 GCGGGACCTCCAGGGGAGGCCGTCTTCTTTCTTCGGCATGGACAGGCCCCCTAAGCTGCG 1085
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         961 GCGGGACCTCCAGGGGAGGCCGTCTTCTTTCTTCGGCATGGACAGGCCCCCTAAGCTGCG 1020

Qy       1086 GGACGAGTTCTGATTCTGGCCGGAGCACCAGACCCTCTGATCCTGGAGGGACCAGAGTCT 1145
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1021 GGACGAGTTCTGATTCTGGCCGGAGCACCAGACCCTCTGATCCTGGAGGGACCAGAGTCT 1080

Qy       1146 GAGCTGGTCCTTCCAGCCAGGCCTGGCAGCCAGAGGTGCTCCGGGATTGCAAACTCCTCT 1205
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1081 GAGCTGGTCCTTCCAGCCAGGCCTGGCAGCCAGAGGTGCTCCGGGATTGCAAACTCCTCT 1140

Qy       1206 TCTCACCTGCCAAAGATGGAGAAGAGTGCCAGGGACCCCTCAAGGAGGGAGACGCTCCAT 1265
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1141 TCTCACCTGCCAAAGATGGAGAAGAGTGCCAGGGACCCCTCAAGGAGGGAGACGCTCCAT 1200

Qy       1266 ATCCCAGGGCATAGGACTTGCAGGTTCCTAGGAGCAGGAGCATCTCCCATCGCACGTGCT 1325
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1201 ATCCCAGGGCATAGGACTTGCAGGTTCCTAGGAGCAGGAGCATCTCCCATCGCACGTGCT 1260

Qy       1326 TTCTGCTCTTCTGGGAATTTCTCACACTGGCAAAGCAGTCCAGCCTCCGTCTTCTGGTCC 1385
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1261 TTCTGCTCTTCTGGGAATTTCTCACACTGGCAAAGCAGTCCAGCCTCCGTCTTCTGGTCC 1320

Qy       1386 ACTCTGCTCTGAGCAGCCTGGGATGCTGAACTCTTCAGAGAGATTTTTTTATAGAGAGAT 1445
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1321 ACTCTGCTCTGAGCAGCCTGGGATGCTGAACTCTTCAGAGAGATTTTTTTATAGAGAGAT 1380

Qy       1446 TTCTATAATTTTGATACAAGGTCATGACTATCCTAGAACTCTCTGTGGTTTTTGAAAATC 1505
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1381 TTCTATAATTTTGATACAAGGTCATGACTATCCTAGAACTCTCTGTGGTTTTTGAAAATC 1440

Qy       1506 ATTGAATTC 1514
              |||||||||
Db        1441 ATTGAATTC 1449
```

```
RESULT 9
BD103191
LOCUS       BD103191                5009 bp    DNA     linear   PAT 27-AUG-2002
DEFINITION  O-fucosyltransferase.
ACCESSION   BD103191
VERSION     BD103191.1  GI:22648765
KEYWORDS    JP 2001527389-A/4.
SOURCE      unidentified
  ORGANISM  unidentified
            unclassified.
REFERENCE   1  (bases 1 to 5009)
  AUTHORS   Wang,Y. and Spellman,M.W.
  TITLE     O-fucosyltransferase
  JOURNAL   Patent: JP 2001527389-A 4 25-DEC-2001;
            GENENTECH INC
COMMENT     OS   Unidentified
            PN   JP 2001527389-A/4
            PD   25-DEC-2001
            PF   17-DEC-1997 JP 1998532877
            PR   31-JAN-1997 US    08/792498,26-NOV-1997 US    08/978741 PI
            YANG WANG,MICHAEL W SPELLMAN
            PC   C12N15/54,C12N9/10,C07K16/40
            CC   Strandedness: Single;
            CC   Topology: Linear;
            CC   O-fucosyltransferase
            FH   Key             Location/Qualifiers
            FT   source          1. .5009
            FT                   /organism='Unidentified'.
FEATURES             Location/Qualifiers
     source          1. .5009
                     /organism="unidentified"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:32644"
ORIGIN
```

```
  Query Match             95.6%;  Score 1447.4;  DB 6;  Length 5009;
  Best Local Similarity   99.9%;  Pred. No. 0;
  Matches 1448;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy         66 GAACCAGGCCGATCACTTCTTGGGCTCTCTGGCATTTGCAAAGCTGCTAAACCGTACCTT 125
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db          1 GAACCAGGCCGATCACTTCTTGGGCTCTCTGGCATTTGCAAAGCTGCTAAACCGTACCTT 60

Qy        126 GGCTGTCCCTCCTTGGATTGAGTACCAGCATCACAAGCCTCCTTTCACCAACCTCCATGT 185
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GGCTGTCCCTCCTTGGATTGAGTACCAGCATCACAAGCCTCCTTTCACCAACCTCCATGT 120

Qy        186 GTCCTACCAGAAGTACTTCAAGCTGGAGCCCCTCCAGGCTTACCATCGGGTCATCAGCTT 245
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 GTCCTACCAGAAGTACTTCAAGCTGGAGCCCCTCCAGGCTTACCATCGGGTCATCAGCTT 180

Qy        246 GGAGGATTTCATGGAGAAGCTGGCACCCACCCACTGGCCCCCTGAGAAGCGGGTGGCATA 305
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 GGAGGATTTCATGGAGAAGCTGGCACCCACCCACTGGCCCCCTGAGAAGCGGGTGGCATA 240

Qy        306 CTGCTTTGAGGTGGCAGCCCAGCGAAGCCCAGATAAGAAGACGTGCCCCATGAAGGAAGG 365
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 CTGCTTTGAGGTGGCAGCCCAGCGAAGCCCAGATAAGAAGACGTGCCCCATGAAGGAAGG 300

Qy        366 AAACCCCTTTGGCCCATTCTGGGATCAGTTTCATGTGAGTTTCAACAAGTCGGAGCTTTT 425
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 AAACCCCTTTGGCCCATTCTGGGATCAGTTTCATGTGAGTTTCAACAAGTCGGAGCTTTT 360

Qy        426 TACAGGCATTTCCTTCAGTGCTTCCTACAGAGAACAATGGAGCCAGAGATTTTCTCCAAA 485
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 TACAGGCATTTCCTTCAGTGCTTCCTACAGAGAACAATGGAGCCAGAGATTTTCTCCAAA 420

Qy        486 GGAACATCCGGTGCTTGCCCTGCCAGGAGCCCCAGCCCAGTTCCCCGTCCTAGAAGAACA 545
              |||||||||||||||||||||||||||||||||||||||||||||||||| |||||
Db        421 GGAACATCCGGTGCTTGCCCTGCCAGGAGCCCCAGCCCAGTTCCCCGTCCTAGAGGAACA 480

Qy        546 CAGGCCACTACAGAAGTACATGGTATGGTCAGACGAAATGGTGAAGACGGGAGAGGCCCA 605
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 CAGGCCACTACAGAAGTACATGGTATGGTCAGACGAAATGGTGAAGACGGGAGAGGCCCA 540

Qy        606 GATTCATGCCCACCTTGTCCGGCCCTATGTGGGCATTCATCTGCGCATTGGCTCTGACTG 665
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 GATTCATGCCCACCTTGTCCGGCCCTATGTGGGCATTCATCTGCGCATTGGCTCTGACTG 600

Qy        666 GAAGAACGCCTGTGCCATGCTGAAGGACGGGACTGCAGGCTCGCACTTCATGGCCTCTCC 725
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 GAAGAACGCCTGTGCCATGCTGAAGGACGGGACTGCAGGCTCGCACTTCATGGCCTCTCC 660

Qy        726 GCAGTGTGTGGGCTACAGCCGCAGCACAGCGGCCCCCCTCACGATGACTATGTGCCTGCC 785
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 GCAGTGTGTGGGCTACAGCCGCAGCACAGCGGCCCCCCTCACGATGACTATGTGCCTGCC 720

Qy        786 TGACCTGAAGGAGATCCAGAGGGCTGTGAAGCTCTGGGTGAGGTCGCTGGATGCCCAGTC 845
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 TGACCTGAAGGAGATCCAGAGGGCTGTGAAGCTCTGGGTGAGGTCGCTGGATGCCCAGTC 780

Qy        846 GGTCTACGTTGCTACTGATTCCGAGAGTTATGTGCCTGAGCTCCAACAGCTCTTCAAAGG 905
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 GGTCTACGTTGCTACTGATTCCGAGAGTTATGTGCCTGAGCTCCAACAGCTCTTCAAAGG 840

Qy        906 GAAGGTGAAGGTGGTGAGCCTGAAGCCTGAGGTGGCCCAGGTCGACCTGTACATCCTCGG 965
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 GAAGGTGAAGGTGGTGAGCCTGAAGCCTGAGGTGGCCCAGGTCGACCTGTACATCCTCGG 900

Qy        966 CCAAGCCGACCACTTTATTGGCAACTGTGTCTCCTCCTTCACTGCCTTTGTGAAGCGGGA 1025
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 CCAAGCCGACCACTTTATTGGCAACTGTGTCTCCTCCTTCACTGCCTTTGTGAAGCGGGA 960

Qy       1026 GCGGGACCTCCAGGGGAGGCCGTCTTCTTTCTTCGGCATGGACAGGCCCCCTAAGCTGCG 1085
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         961 GCGGGACCTCCAGGGGAGGCCGTCTTCTTTCTTCGGCATGGACAGGCCCCCTAAGCTGCG 1020

Qy       1086 GGACGAGTTCTGATTCTGGCCGGAGCACCAGACCCTCTGATCCTGGAGGGACCAGAGTCT 1145
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1021 GGACGAGTTCTGATTCTGGCCGGAGCACCAGACCCTCTGATCCTGGAGGGACCAGAGTCT 1080

Qy       1146 GAGCTGGTCCTTCCAGCCAGGCCTGGCAGCCAGAGGTGCTCCGGGATTGCAAACTCCTCT 1205
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1081 GAGCTGGTCCTTCCAGCCAGGCCTGGCAGCCAGAGGTGCTCCGGGATTGCAAACTCCTCT 1140

Qy       1206 TCTCACCTGCCAAAGATGGAGAAGAGTGCCAGGGACCCCTCAAGGAGGGAGACGCTCCAT 1265
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1141 TCTCACCTGCCAAAGATGGAGAAGAGTGCCAGGGACCCCTCAAGGAGGGAGACGCTCCAT 1200

Qy       1266 ATCCCAGGGCATAGGACTTGCAGGTTCCTAGGAGCAGGAGCATCTCCCATCGCACGTGCT 1325
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1201 ATCCCAGGGCATAGGACTTGCAGGTTCCTAGGAGCAGGAGCATCTCCCATCGCACGTGCT 1260

Qy       1326 TTCTGCTCTTCTGGGAATTTCTCACACTGGCAAAGCAGTCCAGCCTCCGTCTTCTGGTCC 1385
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1261 TTCTGCTCTTCTGGGAATTTCTCACACTGGCAAAGCAGTCCAGCCTCCGTCTTCTGGTCC 1320

Qy       1386 ACTCTGCTCTGAGCAGCCTGGGATGCTGAACTCTTCAGAGAGATTTTTTTATAGAGAGAT 1445
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1321 ACTCTGCTCTGAGCAGCCTGGGATGCTGAACTCTTCAGAGAGATTTTTTTATAGAGAGAT 1380

Qy       1446 TTCTATAATTTTGATACAAGGTCATGACTATCCTAGAACTCTCTGTGGTTTTTGAAAATC 1505
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1381 TTCTATAATTTTGATACAAGGTCATGACTATCCTAGAACTCTCTGTGGTTTTTGAAAATC 1440

Qy       1506 ATTGAATTC 1514
              |||||||||
Db        1441 ATTGAATTC 1449
```

RESULT 10
AR163452

```
LOCUS       AR163452               1300 bp    DNA     linear   PAT 17-OCT-2001
DEFINITION  Sequence 6 from patent US 6270987.
ACCESSION   AR163452
VERSION     AR163452.1  GI:16234057
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 1300)
  AUTHORS   Wang,Y. and Spellman,M.W.
  TITLE     O-fucosyltransferase
  JOURNAL   Patent: US 6270987-A 6 07-AUG-2001;
FEATURES             Location/Qualifiers
     source          1. .1300
                     /organism="unknown"
                     /mol_type="unassigned DNA"
ORIGIN
```

Query Match 76.8%; Score 1163.4; DB 6; Length 1300;
Best Local Similarity 99.9%; Pred. No. 2.7e-284;
Matches 1164; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```
Qy          1 ATGCCCGCGGGCTCCTGGGACCCGGCCGGTTACCTGCTCTACTGCCCCTGCATGGGGCGC 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        136 ATGCCCGCGGGCTCCTGGGACCCGGCCGGTTACCTGCTCTACTGCCCCTGCATGGGGCGC 195

Qy         61 TTTGGGAACCAGGCCGATCACTTCTTGGGCTCTCTGGCATTTGCAAAGCTGCTAAACCGT 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        196 TTTGGGAACCAGGCCGATCACTTCTTGGGCTCTCTGGCATTTGCAAAGCTGCTAAACCGT 255

Qy        121 ACCTTGGCTGTCCCTCCTTGGATTGAGTACCAGCATCACAAGCCTCCTTTCACCAACCTC 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        256 ACCTTGGCTGTCCCTCCTTGGATTGAGTACCAGCATCACAAGCCTCCTTTCACCAACCTC 315

Qy        181 CATGTGTCCTACCAGAAGTACTTCAAGCTGGAGCCCCTCCAGGCTTACCATCGGGTCATC 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        316 CATGTGTCCTACCAGAAGTACTTCAAGCTGGAGCCCCTCCAGGCTTACCATCGGGTCATC 375

Qy        241 AGCTTGGAGGATTTCATGGAGAAGCTGGCACCCACCCACTGGCCCCCTGAGAAGCGGGTG 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        376 AGCTTGGAGGATTTCATGGAGAAGCTGGCACCCACCCACTGGCCCCCTGAGAAGCGGGTG 435

Qy        301 GCATACTGCTTTGAGGTGGCAGCCCAGCGAAGCCCAGATAAGAAGACGTGCCCCATGAAG 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        436 GCATACTGCTTTGAGGTGGCAGCCCAGCGAAGCCCAGATAAGAAGACGTGCCCCATGAAG 495
```

```
Qy        361 GAAGGAAACCCCTTTGGCCCATTCTGGGATCAGTTTCATGTGAGTTTCAACAAGTCGGAG 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        496 GAAGGAAACCCCTTTGGCCCATTCTGGGATCAGTTTCATGTGAGTTTCAACAAGTCGGAG 555

Qy        421 CTTTTTACAGGCATTTCCTTCAGTGCTTCCTACAGAGAACAATGGAGCCAGAGATTTTCT 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        556 CTTTTTACAGGCATTTCCTTCAGTGCTTCCTACAGAGAACAATGGAGCCAGAGATTTTCT 615

Qy        481 CCAAAGGAACATCCGGTGCTTGCCCTGCCAGGAGCCCCAGCCCAGTTCCCCGTCCTAGAA 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        616 CCAAAGGAACATCCGGTGCTTGCCCTGCCAGGAGCCCCAGCCCAGTTCCCCGTCCTAGAG 675

Qy        541 GAACACAGGCCACTACAGAAGTACATGGTATGGTCAGACGAAATGGTGAAGACGGGAGAG 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        676 GAACACAGGCCACTACAGAAGTACATGGTATGGTCAGACGAAATGGTGAAGACGGGAGAG 735

Qy        601 GCCCAGATTCATGCCCACCTTGTCCGGCCCTATGTGGGCATTCATCTGCGCATTGGCTCT 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        736 GCCCAGATTCATGCCCACCTTGTCCGGCCCTATGTGGGCATTCATCTGCGCATTGGCTCT 795

Qy        661 GACTGGAAGAACGCCTGTGCCATGCTGAAGGACGGGACTGCAGGCTCGCACTTCATGGCC 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        796 GACTGGAAGAACGCCTGTGCCATGCTGAAGGACGGGACTGCAGGCTCGCACTTCATGGCC 855

Qy        721 TCTCCGCAGTGTGTGGGCTACAGCCGCAGCACAGCGGCCCCCCTCACGATGACTATGTGC 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        856 TCTCCGCAGTGTGTGGGCTACAGCCGCAGCACAGCGGCCCCCCTCACGATGACTATGTGC 915

Qy        781 CTGCCTGACCTGAAGGAGATCCAGAGGGCTGTGAAGCTCTGGGTGAGGTCGCTGGATGCC 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        916 CTGCCTGACCTGAAGGAGATCCAGAGGGCTGTGAAGCTCTGGGTGAGGTCGCTGGATGCC 975

Qy        841 CAGTCGGTCTACGTTGCTACTGATTCCGAGAGTTATGTGCCTGAGCTCCAACAGCTCTTC 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        976 CAGTCGGTCTACGTTGCTACTGATTCCGAGAGTTATGTGCCTGAGCTCCAACAGCTCTTC 1035

Qy        901 AAAGGGAAGGTGAAGGTGGTGAGCCTGAAGCCTGAGGTGGCCCAGGTCGACCTGTACATC 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1036 AAAGGGAAGGTGAAGGTGGTGAGCCTGAAGCCTGAGGTGGCCCAGGTCGACCTGTACATC 1095

Qy        961 CTCGGCCAAGCCGACCACTTTATTGGCAACTGTGTCTCCTCCTTCACTGCCTTTGTGAAG 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1096 CTCGGCCAAGCCGACCACTTTATTGGCAACTGTGTCTCCTCCTTCACTGCCTTTGTGAAG 1155

Qy       1021 CGGGAGCGGGACCTCCAGGGGAGGCCGTCTTCTTTCTTCGGCATGGACAGGCCCCCTAAG 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1156 CGGGAGCGGGACCTCCAGGGGAGGCCGTCTTCTTTCTTCGGCATGGACAGGCCCCCTAAG 1215

Qy       1081 CTGCGGGACGAGTTCTGATTCTGGCCGGAGCACCAGACCCTCTGATCCTGGAGGGACCAG 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1216 CTGCGGGACGAGTTCTGATTCTGGCCGGAGCACCAGACCCTCTGATCCTGGAGGGACCAG 1275

Qy       1141 AGTCTGAGCTGGTCCTTCCAGCCAG 1165
              |||||||||||||||||||||||||
Db       1276 AGTCTGAGCTGGTCCTTCCAGCCAG 1300
```

```
RESULT 11
BD103189
LOCUS       BD103189               1300 bp    DNA     linear   PAT 27-AUG-2002
DEFINITION  O-fucosyltransferase.
ACCESSION   BD103189
VERSION     BD103189.1  GI:22648763
KEYWORDS    JP 2001527389-A/2.
SOURCE      unidentified
  ORGANISM  unidentified
            unclassified.
REFERENCE   1  (bases 1 to 1300)
  AUTHORS   Wang,Y. and Spellman,M.W.
  TITLE     O-fucosyltransferase
  JOURNAL   Patent: JP 2001527389-A 2 25-DEC-2001;
            GENENTECH INC
COMMENT     OS   Unidentified
            PN   JP 2001527389-A/2
            PD   25-DEC-2001
            PF   17-DEC-1997 JP 1998532877
            PR   31-JAN-1997 US    08/792498,26-NOV-1997 US    08/978741 PI
            YANG WANG,MICHAEL W SPELLMAN
            PC   C12N15/54,C12N9/10,C07K16/40
            CC   Strandedness: Single;
            CC   Topology: Linear;
            CC   O-fucosyltransferase
            FH   Key             Location/Qualifiers
            FT   source          1. .1300
            FT                   /organism='Unidentified'.
FEATURES             Location/Qualifiers
     source          1. .1300
                     /organism="unidentified"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:32644"
ORIGIN
```

```
  Query Match             76.8%;  Score 1163.4;  DB 6;  Length 1300;
  Best Local Similarity   99.9%;  Pred. No. 2.7e-284;
  Matches 1164;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 ATGCCCGCGGGCTCCTGGGACCCGGCCGGTTACCTGCTCTACTGCCCCTGCATGGGGCGC 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        136 ATGCCCGCGGGCTCCTGGGACCCGGCCGGTTACCTGCTCTACTGCCCCTGCATGGGGCGC 195

Qy         61 TTTGGGAACCAGGCCGATCACTTCTTGGGCTCTCTGGCATTTGCAAAGCTGCTAAACCGT 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        196 TTTGGGAACCAGGCCGATCACTTCTTGGGCTCTCTGGCATTTGCAAAGCTGCTAAACCGT 255

Qy        121 ACCTTGGCTGTCCCTCCTTGGATTGAGTACCAGCATCACAAGCCTCCTTTCACCAACCTC 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        256 ACCTTGGCTGTCCCTCCTTGGATTGAGTACCAGCATCACAAGCCTCCTTTCACCAACCTC 315

Qy        181 CATGTGTCCTACCAGAAGTACTTCAAGCTGGAGCCCCTCCAGGCTTACCATCGGGTCATC 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        316 CATGTGTCCTACCAGAAGTACTTCAAGCTGGAGCCCCTCCAGGCTTACCATCGGGTCATC 375

Qy        241 AGCTTGGAGGATTTCATGGAGAAGCTGGCACCCACCCACTGGCCCCCTGAGAAGCGGGTG 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        376 AGCTTGGAGGATTTCATGGAGAAGCTGGCACCCACCCACTGGCCCCCTGAGAAGCGGGTG 435

Qy        301 GCATACTGCTTTGAGGTGGCAGCCCAGCGAAGCCCAGATAAGAAGACGTGCCCCATGAAG 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        436 GCATACTGCTTTGAGGTGGCAGCCCAGCGAAGCCCAGATAAGAAGACGTGCCCCATGAAG 495

Qy        361 GAAGGAAACCCCTTTGGCCCATTCTGGGATCAGTTTCATGTGAGTTTCAACAAGTCGGAG 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        496 GAAGGAAACCCCTTTGGCCCATTCTGGGATCAGTTTCATGTGAGTTTCAACAAGTCGGAG 555

Qy        421 CTTTTTACAGGCATTTCCTTCAGTGCTTCCTACAGAGAACAATGGAGCCAGAGATTTTCT 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        556 CTTTTTACAGGCATTTCCTTCAGTGCTTCCTACAGAGAACAATGGAGCCAGAGATTTTCT 615

Qy        481 CCAAAGGAACATCCGGTGCTTGCCCTGCCAGGAGCCCCAGCCCAGTTCCCCGTCCTAGAA 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        616 CCAAAGGAACATCCGGTGCTTGCCCTGCCAGGAGCCCCAGCCCAGTTCCCCGTCCTAGAG 675

Qy        541 GAACACAGGCCACTACAGAAGTACATGGTATGGTCAGACGAAATGGTGAAGACGGGAGAG 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        676 GAACACAGGCCACTACAGAAGTACATGGTATGGTCAGACGAAATGGTGAAGACGGGAGAG 735

Qy        601 GCCCAGATTCATGCCCACCTTGTCCGGCCCTATGTGGGCATTCATCTGCGCATTGGCTCT 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        736 GCCCAGATTCATGCCCACCTTGTCCGGCCCTATGTGGGCATTCATCTGCGCATTGGCTCT 795

Qy        661 GACTGGAAGAACGCCTGTGCCATGCTGAAGGACGGGACTGCAGGCTCGCACTTCATGGCC 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        796 GACTGGAAGAACGCCTGTGCCATGCTGAAGGACGGGACTGCAGGCTCGCACTTCATGGCC 855

Qy        721 TCTCCGCAGTGTGTGGGCTACAGCCGCAGCACAGCGGCCCCCCTCACGATGACTATGTGC 780
```

```
Db        856 TCTCCGCAGTGTGTGGGCTACAGCCGCAGCACAGCGGCCCCCCTCACGATGACTATGTGC 915

Qy        781 CTGCCTGACCTGAAGGAGATCCAGAGGGCTGTGAAGCTCTGGGTGAGGTCGCTGGATGCC 840

Db        916 CTGCCTGACCTGAAGGAGATCCAGAGGGCTGTGAAGCTCTGGGTGAGGTCGCTGGATGCC 975

Qy        841 CAGTCGGTCTACGTTGCTACTGATTCCGAGAGTTATGTGCCTGAGCTCCAACAGCTCTTC 900

Db        976 CAGTCGGTCTACGTTGCTACTGATTCCGAGAGTTATGTGCCTGAGCTCCAACAGCTCTTC 1035

Qy        901 AAAGGGAAGGTGAAGGTGGTGAGCCTGAAGCCTGAGGTGGCCCAGGTCGACCTGTACATC 960

Db       1036 AAAGGGAAGGTGAAGGTGGTGAGCCTGAAGCCTGAGGTGGCCCAGGTCGACCTGTACATC 1095

Qy        961 CTCGGCCAAGCCGACCACTTTATTGGCAACTGTGTCTCCTCCTTCACTGCCTTTGTGAAG 1020

Db       1096 CTCGGCCAAGCCGACCACTTTATTGGCAACTGTGTCTCCTCCTTCACTGCCTTTGTGAAG 1155

Qy       1021 CGGGAGCGGGACCTCCAGGGGAGGCCGTCTTCTTTCTTCGGCATGGACAGGCCCCCTAAG 1080

Db       1156 CGGGAGCGGGACCTCCAGGGGAGGCCGTCTTCTTTCTTCGGCATGGACAGGCCCCCTAAG 1215

Qy       1081 CTGCGGGACGAGTTCTGATTCTGGCCGGAGCACCAGACCCTCTGATCCTGGAGGGACCAG 1140

Db       1216 CTGCGGGACGAGTTCTGATTCTGGCCGGAGCACCAGACCCTCTGATCCTGGAGGGACCAG 1275

Qy       1141 AGTCTGAGCTGGTCCTTCCAGCCAG 1165

Db       1276 AGTCTGAGCTGGTCCTTCCAGCCAG 1300


RESULT 12
AR163457
LOCUS       AR163457               1100 bp    DNA     linear   PAT 17-OCT-2001
DEFINITION  Sequence 16 from patent US 6270987.
ACCESSION   AR163457
VERSION     AR163457.1  GI:16234062
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 1100)
  AUTHORS   Wang,Y. and Spellman,M.W.
  TITLE     O-fucosyltransferase
  JOURNAL   Patent: US 6270987-A 16 07-AUG-2001;
FEATURES             Location/Qualifiers
     source          1. .1100
                     /organism="unknown"
                     /mol_type="unassigned DNA"
ORIGIN

  Query Match           72.7%;  Score 1100;  DB 6;  Length 1100;
  Best Local Similarity   100.0%;  Pred. No. 3.4e-268;
  Matches 1100;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 ATGCCCGCGGGCTCCTGGGACCCGGCCGGTTACCTGCTCTACTGCCCCTGCATGGGGCGC 60

Db          1 ATGCCCGCGGGCTCCTGGGACCCGGCCGGTTACCTGCTCTACTGCCCCTGCATGGGGCGC 60

Qy         61 TTTGGGAACCAGGCCGATCACTTCTTGGGCTCTCTGGCATTTGCAAAGCTGCTAAACCGT 120

Db         61 TTTGGGAACCAGGCCGATCACTTCTTGGGCTCTCTGGCATTTGCAAAGCTGCTAAACCGT 120

Qy        121 ACCTTGGCTGTCCCTCCTTGGATTGAGTACCAGCATCACAAGCCTCCTTTCACCAACCTC 180

Db        121 ACCTTGGCTGTCCCTCCTTGGATTGAGTACCAGCATCACAAGCCTCCTTTCACCAACCTC 180

Qy        181 CATGTGTCCTACCAGAAGTACTTCAAGCTGGAGCCCCTCCAGGCTTACCATCGGGTCATC 240

Db        181 CATGTGTCCTACCAGAAGTACTTCAAGCTGGAGCCCCTCCAGGCTTACCATCGGGTCATC 240

Qy        241 AGCTTGGAGGATTTCATGGAGAAGCTGGCACCCACCCACTGGCCCCCTGAGAAGCGGGTG 300

Db        241 AGCTTGGAGGATTTCATGGAGAAGCTGGCACCCACCCACTGGCCCCCTGAGAAGCGGGTG 300

Qy        301 GCATACTGCTTTGAGGTGGCAGCCCAGCGAAGCCCAGATAAGAAGACGTGCCCCATGAAG 360

Db        301 GCATACTGCTTTGAGGTGGCAGCCCAGCGAAGCCCAGATAAGAAGACGTGCCCCATGAAG 360

Qy        361 GAAGGAAACCCCCTTTGGCCCATTCTGGGATCAGTTTCATGTGAGTTTCAACAAGTCGGAG 420

Db        361 GAAGGAAACCCCCTTTGGCCCATTCTGGGATCAGTTTCATGTGAGTTTCAACAAGTCGGAG 420

Qy        421 CTTTTTACAGGCATTTCCTTCAGTGCTTCCTACAGAGAACAATGGAGCCAGAGATTTTCT 480

Db        421 CTTTTTACAGGCATTTCCTTCAGTGCTTCCTACAGAGAACAATGGAGCCAGAGATTTTCT 480

Qy        481 CCAAAGGAACATCCGGTGCTTGCCCTGCCAGGAGCCCCAGCCCAGTTCCCCGTCCTAGAA 540

Db        481 CCAAAGGAACATCCGGTGCTTGCCCTGCCAGGAGCCCCAGCCCAGTTCCCCGTCCTAGAA 540

Qy        541 GAACACAGGCCACTACAGAAGTACATGGTATGGTCAGACGAAATGGTGAAGACGGGAGAG 600

Db        541 GAACACAGGCCACTACAGAAGTACATGGTATGGTCAGACGAAATGGTGAAGACGGGAGAG 600

Qy        601 GCCCAGATTCATGCCCACCTTGTCCGGCCCTATGTGGGCATTCATCTGCGCATTGGCTCT 660

Db        601 GCCCAGATTCATGCCCACCTTGTCCGGCCCTATGTGGGCATTCATCTGCGCATTGGCTCT 660

Qy        661 GACTGGAAGAACGCCTGTGCCATGCTGAAGGACGGGACTGCAGGCTCGCACTTCATGGCC 720

Db        661 GACTGGAAGAACGCCTGTGCCATGCTGAAGGACGGGACTGCAGGCTCGCACTTCATGGCC 720

Qy        721 TCTCCGCAGTGTGTGGGCTACAGCCGCAGCACAGCGGCCCCCCTCACGATGACTATGTGC 780

Db        721 TCTCCGCAGTGTGTGGGCTACAGCCGCAGCACAGCGGCCCCCCTCACGATGACTATGTGC 780

Qy        781 CTGCCTGACCTGAAGGAGATCCAGAGGGCTGTGAAGCTCTGGGTGAGGTCGCTGGATGCC 840

Db        781 CTGCCTGACCTGAAGGAGATCCAGAGGGCTGTGAAGCTCTGGGTGAGGTCGCTGGATGCC 840

Qy        841 CAGTCGGTCTACGTTGCTACTGATTCCGAGAGTTATGTGCCTGAGCTCCAACAGCTCTTC 900

Db        841 CAGTCGGTCTACGTTGCTACTGATTCCGAGAGTTATGTGCCTGAGCTCCAACAGCTCTTC 900

Qy        901 AAAGGGAAGGTGAAGGTGGTGAGCCTGAAGCCTGAGGTGGCCCAGGTCGACCTGTACATC 960

Db        901 AAAGGGAAGGTGAAGGTGGTGAGCCTGAAGCCTGAGGTGGCCCAGGTCGACCTGTACATC 960

Qy        961 CTCGGCCAAGCCGACCACTTTATTGGCAACTGTGTCTCCTCCTTCACTGCCTTTGTGAAG 1020

Db        961 CTCGGCCAAGCCGACCACTTTATTGGCAACTGTGTCTCCTCCTTCACTGCCTTTGTGAAG 1020

Qy       1021 CGGGAGCGGGACCTCCAGGGGAGGCCGTCTTCTTTCTTCGGCATGGACAGGCCCCCTAAG 1080

Db       1021 CGGGAGCGGGACCTCCAGGGGAGGCCGTCTTCTTTCTTCGGCATGGACAGGCCCCCTAAG 1080

Qy       1081 CTGCGGGACGAGTTCTGATT 1100

Db       1081 CTGCGGGACGAGTTCTGATT 1100


RESULT 13
BD103196
LOCUS       BD103196               1100 bp    DNA     linear   PAT 27-AUG-2002
DEFINITION  O-fucosyltransferase.
ACCESSION   BD103196
VERSION     BD103196.1  GI:22648770
KEYWORDS    JP 2001527389-A/9.
SOURCE      unidentified
  ORGANISM  unidentified
            unclassified.
REFERENCE   1  (bases 1 to 1100)
  AUTHORS   Wang,Y. and Spellman,M.W.
  TITLE     O-fucosyltransferase
```

```
  JOURNAL    Patent: JP 2001527389-A 9 25-DEC-2001;
             GENENTECH INC
COMMENT      OS   Unidentified
             PN   JP 2001527389-A/9
             PD   25-DEC-2001
             PF   17-DEC-1997 JP 1998532877
             PR   31-JAN-1997 US    08/792498,26-NOV-1997 US    08/978741 PI
             YANG WANG,MICHAEL W SPELLMAN
             PC   C12N15/54,C12N9/10,C07K16/40
             CC   Strandedness: Single;
             CC   Topology: Linear;
             CC   O-fucosyltransferase
             FH   Key             Location/Qualifiers
             FT   source          1. .1100
             FT                   /organism='Unidentified'.
FEATURES               Location/Qualifiers
     source            1. .1100
                       /organism="unidentified"
                       /mol_type="genomic DNA"
                       /db_xref="taxon:32644"
ORIGIN

  Query Match             72.7%;  Score 1100;  DB 6;  Length 1100;
  Best Local Similarity   100.0%;  Pred. No. 3.4e-268;
  Matches 1100;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 ATGCCCGCGGGCTCCTGGGACCCGGCCGGTTACCTGCTCTACTGCCCCTGCATGGGGCGC 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGCCCGCGGGCTCCTGGGACCCGGCCGGTTACCTGCTCTACTGCCCCTGCATGGGGCGC 60

Qy         61 TTTGGGAACCAGGCCGATCACTTCTTGGGCTCTCTGGCATTTGCAAAGCTGCTAAACCGT 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 TTTGGGAACCAGGCCGATCACTTCTTGGGCTCTCTGGCATTTGCAAAGCTGCTAAACCGT 120

Qy        121 ACCTTGGCTGTCCCTCCTTGGATTGAGTACCAGCATCACAAGCCTCCTTTCACCAACCTC 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 ACCTTGGCTGTCCCTCCTTGGATTGAGTACCAGCATCACAAGCCTCCTTTCACCAACCTC 180

Qy        181 CATGTGTCCTACCAGAAGTACTTCAAGCTGGAGCCCCTCCAGGCTTACCATCGGGTCATC 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 CATGTGTCCTACCAGAAGTACTTCAAGCTGGAGCCCCTCCAGGCTTACCATCGGGTCATC 240

Qy        241 AGCTTGGAGGATTTCATGGAGAAGCTGGCACCCACCCACTGGCCCCCTGAGAAGCGGGTG 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 AGCTTGGAGGATTTCATGGAGAAGCTGGCACCCACCCACTGGCCCCCTGAGAAGCGGGTG 300

Qy        301 GCATACTGCTTTGAGGTGGCAGCCCAGCGAAGCCCAGATAAGAAGACGTGCCCCATGAAG 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 GCATACTGCTTTGAGGTGGCAGCCCAGCGAAGCCCAGATAAGAAGACGTGCCCCATGAAG 360

Qy        361 GAAGGAAACCCCTTTGGCCCATTCTGGGATCAGTTTCATGTGAGTTTCAACAAGTCGGAG 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 GAAGGAAACCCCTTTGGCCCATTCTGGGATCAGTTTCATGTGAGTTTCAACAAGTCGGAG 420

Qy        421 CTTTTTACAGGCATTTCCTTCAGTGCTTCCTACAGAGAACAATGGAGCCAGAGATTTTCT 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 CTTTTTACAGGCATTTCCTTCAGTGCTTCCTACAGAGAACAATGGAGCCAGAGATTTTCT 480

Qy        481 CCAAAGGAACATCCGGTGCTTGCCCTGCCAGGAGCCCCAGCCCAGTTCCCCGTCCTAGAA 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 CCAAAGGAACATCCGGTGCTTGCCCTGCCAGGAGCCCCAGCCCAGTTCCCCGTCCTAGAA 540

Qy        541 GAACACAGGCCACTACAGAAGTACATGGTATGGTCAGACGAAATGGTGAAGACGGGAGAG 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 GAACACAGGCCACTACAGAAGTACATGGTATGGTCAGACGAAATGGTGAAGACGGGAGAG 600

Qy        601 GCCCAGATTCATGCCCACCTTGTCCGGCCCTATGTGGGCATTCATCTGCGCATTGGCTCT 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 GCCCAGATTCATGCCCACCTTGTCCGGCCCTATGTGGGCATTCATCTGCGCATTGGCTCT 660

Qy        661 GACTGGAAGAACGCCTGTGCCATGCTGAAGGACGGGACTGCAGGCTCGCACTTCATGGCC 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 GACTGGAAGAACGCCTGTGCCATGCTGAAGGACGGGACTGCAGGCTCGCACTTCATGGCC 720

Qy        721 TCTCCGCAGTGTGTGGGCTACAGCCGCAGCACAGCGGCCCCCCTCACGATGACTATGTGC 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 TCTCCGCAGTGTGTGGGCTACAGCCGCAGCACAGCGGCCCCCCTCACGATGACTATGTGC 780

Qy        781 CTGCCTGACCTGAAGGAGATCCAGAGGGCTGTGAAGCTCTGGGTGAGGTCGCTGGATGCC 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 CTGCCTGACCTGAAGGAGATCCAGAGGGCTGTGAAGCTCTGGGTGAGGTCGCTGGATGCC 840

Qy        841 CAGTCGGTCTACGTTGCTACTGATTCCGAGAGTTATGTGCCTGAGCTCCAACAGCTCTTC 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 CAGTCGGTCTACGTTGCTACTGATTCCGAGAGTTATGTGCCTGAGCTCCAACAGCTCTTC 900

Qy        901 AAAGGGAAGGTGAAGGTGGTGAGCCTGAAGCCTGAGGTGGCCCAGGTCGACCTGTACATC 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 AAAGGGAAGGTGAAGGTGGTGAGCCTGAAGCCTGAGGTGGCCCAGGTCGACCTGTACATC 960

Qy        961 CTCGGCCAAGCCGACCACTTTATTGGCAACTGTGTCTCCTCCTTCACTGCCTTTGTGAAG 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 CTCGGCCAAGCCGACCACTTTATTGGCAACTGTGTCTCCTCCTTCACTGCCTTTGTGAAG 1020

Qy       1021 CGGGAGCGGGACCTCCAGGGGAGGCCGTCTTCTTTCTTCGGCATGGACAGGCCCCCTAAG 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1021 CGGGAGCGGGACCTCCAGGGGAGGCCGTCTTCTTTCTTCGGCATGGACAGGCCCCCTAAG 1080

Qy       1081 CTGCGGGACGAGTTCTGATT 1100
              ||||||||||||||||||||
Db       1081 CTGCGGGACGAGTTCTGATT 1100


RESULT 14
CQ727777
LOCUS       CQ727777               1167 bp    DNA     linear   PAT 03-FEB-2004
DEFINITION  Sequence 13711 from Patent WO02068579.
ACCESSION   CQ727777
VERSION     CQ727777.1  GI:42294771
KEYWORDS    .
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1
  AUTHORS   Venter,C.J., Adams,M.C., Li,P.W. and Myers,E.W.
  TITLE     Kits, such as nucleic acid arrays, comprising a majority of
            humanexons or transcripts, for detecting expression and other uses
            thereof
  JOURNAL   Patent: WO 02068579-A 13711 06-SEP-2002;
            PE Corporation (NY) (US)
FEATURES               Location/Qualifiers
     source            1. .1167
                       /organism="Homo sapiens"
                       /mol_type="unassigned DNA"
                       /db_xref="taxon:9606"
ORIGIN

  Query Match             72.4%;  Score 1096.4;  DB 6;  Length 1167;
  Best Local Similarity   99.9%;  Pred. No. 2.8e-267;
  Matches 1097;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 ATGCCCGCGGGCTCCTGGGACCCGGCCGGTTACCTGCTCTACTGCCCCTGCATGGGGCGC 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         70 ATGCCCGCGGGCTCCTGGGACCCGGCCGGTTACCTGCTCTACTGCCCCTGCATGGGGCGC 129

Qy         61 TTTGGGAACCAGGCCGATCACTTCTTGGGCTCTCTGGCATTTGCAAAGCTGCTAAACCGT 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        130 TTTGGGAACCAGGCCGATCACTTCTTGGGCTCTCTGGCATTTGCAAAGCTGCTAAACCGT 189

Qy        121 ACCTTGGCTGTCCCTCCTTGGATTGAGTACCAGCATCACAAGCCTCCTTTCACCAACCTC 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        190 ACCTTGGCTGTCCCTCCTTGGATTGAGTACCAGCATCACAAGCCTCCTTTCACCAACCTC 249

Qy        181 CATGTGTCCTACCAGAAGTACTTCAAGCTGGAGCCCCTCCAGGCTTACCATCGGGTCATC 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db        250 CATGTGTCCTACCAGAAGTACTTCAAGCTGGAGCCCCTCCAGGCTTACCATCGGGTCATC 309

Qy        241 AGCTTGGAGGATTTCATGGAGAAGCTGGCACCCACCCACTGGCCCCCTGAGAAGCGGGTG 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        310 AGCTTGGAGGATTTCATGGAGAAGCTGGCACCCACCCACTGGCCCCCTGAGAAGCGGGTG 369

Qy        301 GCATACTGCTTTGAGGTGGCAGCCCAGCGAAGCCCAGATAAGAAGACGTGCCCCATGAAG 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        370 GCATACTGCTTTGAGGTGGCAGCCCAGCGAAGCCCAGATAAGAAGACGTGCCCCATGAAG 429

Qy        361 GAAGGAAACCCCTTTGGCCCATTCTGGGATCAGTTTCATGTGAGTTTCAACAAGTCGGAG 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        430 GAAGGAAACCCCTTTGGCCCATTCTGGGATCAGTTTCATGTGAGTTTCAACAAGTCGGAG 489

Qy        421 CTTTTTACAGGCATTTCCTTCAGTGCTTCCTACAGAGAACAATGGAGCCAGAGATTTTCT 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        490 CTTTTTACAGGCATTTCCTTCAGTGCTTCCTACAGAGAACAATGGAGCCAGAGATTTTCT 549

Qy        481 CCAAAGGAACATCCGGTGCTTGCCCTGCCAGGAGCCCCAGCCCAGTTCCCCGTCCTAGAA 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        550 CCAAAGGAACATCCGGTGCTTGCCCTGCCAGGAGCCCCAGCCCAGTTCCCCGTCCTAGAG 609

Qy        541 GAACACAGGCCACTACAGAAGTACATGGTATGGTCAGACGAAATGGTGAAGACGGGAGAG 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        610 GAACACAGGCCACTACAGAAGTACATGGTATGGTCAGACGAAATGGTGAAGACGGGAGAG 669

Qy        601 GCCCAGATTCATGCCCACCTTGTCCGGCCCTATGTGGGCATTCATCTGCGCATTGGCTCT 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        670 GCCCAGATTCATGCCCACCTTGTCCGGCCCTATGTGGGCATTCATCTGCGCATTGGCTCT 729

Qy        661 GACTGGAAGAACGCCTGTGCCATGCTGAAGGACGGGACTGCAGGCTCGCACTTCATGGCC 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        730 GACTGGAAGAACGCCTGTGCCATGCTGAAGGACGGGACTGCAGGCTCGCACTTCATGGCC 789

Qy        721 TCTCCGCAGTGTGTGGGCTACAGCCGCAGCACAGCGGCCCCCCTCACGATGACTATGTGC 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        790 TCTCCGCAGTGTGTGGGCTACAGCCGCAGCACAGCGGCCCCCCTCACGATGACTATGTGC 849

Qy        781 CTGCCTGACCTGAAGGAGATCCAGAGGGCTGTGAAGCTCTGGGTGAGGTCGCTGGATGCC 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        850 CTGCCTGACCTGAAGGAGATCCAGAGGGCTGTGAAGCTCTGGGTGAGGTCGCTGGATGCC 909

Qy        841 CAGTCGGTCTACGTTGCTACTGATTCCGAGAGTTATGTGCCTGAGCTCCAACAGCTCTTC 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        910 CAGTCGGTCTACGTTGCTACTGATTCCGAGAGTTATGTGCCTGAGCTCCAACAGCTCTTC 969

Qy        901 AAAGGGAAGGTGAAGGTGGTGAGCCTGAAGCCTGAGGTGGCCCAGGTCGACCTGTACATC 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        970 AAAGGGAAGGTGAAGGTGGTGAGCCTGAAGCCTGAGGTGGCCCAGGTCGACCTGTACATC 1029

Qy        961 CTCGGCCAAGCCGACCACTTTATTGGCAACTGTGTCTCCTCCTTCACTGCCTTTGTGAAG 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1030 CTCGGCCAAGCCGACCACTTTATTGGCAACTGTGTCTCCTCCTTCACTGCCTTTGTGAAG 1089

Qy       1021 CGGGAGCGGGACCTCCAGGGGAGGCCGTCTTCTTTCTTCGGCATGGACAGGCCCCCTAAG 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1090 CGGGAGCGGGACCTCCAGGGGAGGCCGTCTTCTTTCTTCGGCATGGACAGGCCCCCTAAG 1149

Qy       1081 CTGCGGGACGAGTTCTGA 1098
              ||||||||||||||||||
Db       1150 CTGCGGGACGAGTTCTGA 1167


RESULT 15
AJ781500
LOCUS       AJ781500               1167 bp    mRNA    linear   PRI 07-JUL-2004
DEFINITION  Pan troglodytes mRNA for protein-O-fucosyltransferase 1 (fut12
            gene).
ACCESSION   AJ781500
VERSION     AJ781500.1  GI:50057069
KEYWORDS    fut12 gene; pofut1 gene; protein-O-fucosyltransferase 1.
SOURCE      Pan troglodytes (chimpanzee)
  ORGANISM  Pan troglodytes
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Pan.
REFERENCE   1
  AUTHORS   Martinez-Duncker,I., Mollicone,R., Candelier,J.J., Breton,C. and
            Oriol,R.
  TITLE     A new superfamily of protein-O-fucosyltransferases,
            alpha2-fucosyltransferases, and alpha6-fucosyltransferases:
            phylogeny and identification of conserved peptide motifs
  JOURNAL   Glycobiology 13 (12), 1C-5C (2003)
   PUBMED   12966037
REFERENCE   2
  AUTHORS   Martinez-Duncker,I., Oriol,R. and Mollicone,R.
  TITLE     Phylogeny of fucosyltransferases
  JOURNAL   Unpublished
REFERENCE   3  (bases 1 to 1167)
  AUTHORS   Oriol,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (06-JUL-2004) Oriol R., U504, Inserm, 16 Av. Paul
            Vaillant-Couturier, 94807, FRANCE
FEATURES             Location/Qualifiers
     source          1. .1167
                     /organism="Pan troglodytes"
                     /mol_type="mRNA"
                     /db_xref="taxon:9598"
     gene            1. .1167
                     /gene="fut12"
     CDS             1. .1167
                     /gene="fut12"
                     /function="adds fucose to Ser or Thr"
                     /note="pofut1"
                     /codon_start=1
                     /product="protein-O-fucosyltransferase 1"
                     /protein_id="CAH03712.1"
                     /db_xref="GI:50057070"
                     /translation="MGAAAWARPLSVSFLLLLLPLPGMPAGSWDPAGYLLYCPCMGRF
                     GNQADHFLGSLAFAKLLNRTLAVPPWIEYQHHKPPFTNLHVSYQKYFKLEPLQAYHRV
                     ISLEDFMEKLAPTHWPPEKRVAYCFEVAAQRSPDKKTCPMKEGNPFGPFWDQFHVSFN
                     KSELFTGISFSASYREQWSQRFSPKEHPVLALPGAPAQFPVLEEHRPLQKYMVWSDEM
                     VKTGEAQIHAHLVRPYVGIHLRIGSDWKNACAMLKDGTAGSHFMASPQCVGYSRSTAA
                     PLTMTMCLPDLKEIQRAVKLWVRSLDAQSVYVATDSESYVPELQQLFKGKVKVVSLKP
                     EVAQVDLYILGQADHFIGNCVSSFTAFVKRERDLQGRPSSFFGMDRPPKLRDEF"
ORIGIN

  Query Match             71.8%;  Score 1086.8;  DB 9;  Length 1167;
  Best Local Similarity   99.4%;  Pred. No. 7.7e-265;
  Matches 1091;  Conservative    0;  Mismatches    7;  Indels    0;  Gaps    0;

Qy          1 ATGCCCGCGGGCTCCTGGGACCCGGCCGGTTACCTGCTCTACTGCCCCTGCATGGGGCGC 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         70 ATGCCCGCGGGCTCCTGGGACCCGGCCGGTTACCTGCTCTACTGCCCCTGCATGGGGCGC 129

Qy         61 TTTGGGAACCAGGCCGATCACTTCTTGGGCTCTCTGGCATTTGCAAAGCTGCTAAACCGT 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        130 TTTGGGAACCAGGCCGATCACTTCTTGGGCTCTCTGGCATTTGCAAAGCTGCTAAACCGT 189

Qy        121 ACCTTGGCTGTCCCTCCTTGGATTGAGTACCAGCATCACAAGCCTCCTTTCACCAACCTC 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        190 ACCTTGGCTGTCCCTCCTTGGATTGAGTACCAGCATCACAAGCCTCCTTTCACCAACCTC 249

Qy        181 CATGTGTCCTACCAGAAGTACTTCAAGCTGGAGCCCCTCCAGGCTTACCATCGGGTCATC 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        250 CATGTGTCCTACCAGAAGTACTTCAAGCTGGAGCCCCTCCAGGCTTACCATCGGGTCATC 309

Qy        241 AGCTTGGAGGATTTCATGGAGAAGCTGGCACCCACCCACTGGCCCCCTGAGAAGCGGGTG 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        310 AGCTTGGAGGATTTCATGGAGAAGCTGGCACCCACCCACTGGCCCCCTGAGAAGCGGGTG 369

Qy        301 GCATACTGCTTTGAGGTGGCAGCCCAGCGAAGCCCAGATAAGAAGACGTGCCCCATGAAG 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        370 GCATACTGCTTTGAGGTGGCAGCCCAGCGAAGCCCAGATAAGAAGACGTGCCCCATGAAG 429

Qy        361 GAAGGAAACCCCTTTGGCCCATTCTGGGATCAGTTTCATGTGAGTTTCAACAAGTCGGAG 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db        430 GAAGGAAACCCCTTTGGCCCATTCTGGGATCAGTTTCATGTGAGTTTCAACAAGTCGGAG 489

Qy        421 CTTTTTACAGGCATTTCCTTCAGTGCTTCCTACAGAGAACAATGGAGCCAGAGATTTTCT 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        490 CTTTTTACAGGCATTTCCTTCAGTGCTTCCTACAGAGAACAATGGAGCCAGAGATTTTCT 549

Qy        481 CCAAAGGAACATCCGGTGCTTGCCCTGCCAGGAGCCCCAGCCCAGTTCCCCGTCCTAGAA 540
              |||||||||||||||||||||||||||||||||||| |||||||||||||||| ||
Db        550 CCAAAGGAACATCCGGTGCTTGCCCTGCCAGGAGCCCCGGCCCAGTTCCCCGTCCTGGAG 609

Qy        541 GAACACAGGCCACTACAGAAGTACATGGTATGGTCAGACGAAATGGTGAAGACGGGAGAG 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        610 GAACACAGGCCACTACAGAAGTACATGGTATGGTCAGACGAAATGGTGAAGACGGGAGAG 669

Qy        601 GCCCAGATTCATGCCCACCTTGTCCGGCCCTATGTGGGCATTCATCTGCGCATTGGCTCT 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        670 GCCCAGATTCATGCCCACCTTGTCCGGCCCTATGTGGGCATTCATCTGCGCATTGGCTCT 729

Qy        661 GACTGGAAGAACGCCTGTGCCATGCTGAAGGACGGGACTGCAGGCTCGCACTTCATGGCC 720
              |||||||||||||||||||||||||||||||||||| ||||||||||||||||||||||
Db        730 GACTGGAAGAACGCCTGTGCCATGCTGAAGGACGGGACCGCAGGCTCGCACTTCATGGCC 789

Qy        721 TCTCCGCAGTGTGTGGGCTACAGCCGCAGCACAGCGGCCCCCCTCACGATGACTATGTGC 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        790 TCTCCGCAGTGTGTGGGCTACAGCCGCAGCACAGCGGCCCCCCTCACGATGACTATGTGC 849

Qy        781 CTGCCTGACCTGAAGGAGATCCAGAGGGCTGTGAAGCTCTGGGTGAGGTCGCTGGATGCC 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        850 CTGCCTGACCTGAAGGAGATCCAGAGGGCTGTGAAGCTCTGGGTGAGGTCGCTGGATGCC 909

Qy        841 CAGTCGGTCTACGTTGCTACTGATTCCGAGAGTTATGTGCCTGAGCTCCAACAGCTCTTC 900
              ||||||||||| |||||||||||||||||||||||||||||||||||||||||||||||
Db        910 CAGTCGGTCTATGTTGCTACTGATTCCGAGAGTTATGTGCCTGAGCTCCAACAGCTCTTC 969

Qy        901 AAAGGGAAGGTGAAGGTGGTGAGCCTGAAGCCTGAGGTGGCCCAGGTCGACCTGTACATC 960
              |||||||||||||||||||||||||||||||||||||||||||| ||||||||||||
Db        970 AAAGGGAAGGTGAAGGTGGTGAGCCTGAAGCCTGAGGTGGCCCAGGTTGACCTGTACATC 1029

Qy        961 CTCGGCCAAGCCGACCACTTTATTGGCAACTGTGTCTCCTCCTTCACTGCCTTTGTGAAG 1020
              |||||||||||| ||||||||||||||||||||||||||||||||||||||||||||||
Db       1030 CTCGGCCAAGCTGACCACTTTATTGGCAACTGTGTCTCCTCCTTCACTGCCTTTGTGAAG 1089

Qy       1021 CGGGAGCGGGACCTCCAGGGGAGGCCGTCTTCTTTCTTCGGCATGGACAGGCCCCCTAAG 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1090 CGGGAGCGGGACCTCCAGGGGAGGCCGTCTTCTTTCTTCGGCATGGACAGGCCCCCTAAG 1149

Qy       1081 CTGCGGGACGAGTTCTGA 1098
              ||||||||||||||||||
Db       1150 CTGCGGGACGAGTTCTGA 1167
```

Search completed: November 9, 2005, 07:59:57
Job time : 6834.38 secs

GenCore version 5.1.6

OM nucleic - nucleic search, using sw model

Run on: November 8, 2005, 23:23:13 ; Search time 883.842 Seconds
(without alignments)
10140.369 Million cell updates/sec

Title: US-09-774-954-1
Perfect score: 1514
Sequence: 1 ATGCCCGCGGGCTCCTGGGA.....TTTTGAAAATCATTGAATTC 1514

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 4390206 seqs, 2959870667 residues

Total number of hits satisfying chosen parameters: 8780412

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_16Dec04:*
1: geneseqn1980s:*
2: geneseqn1990s:*
3: geneseqn2000s:*
4: geneseqn2001as:*
5: geneseqn2001bs:*
6: geneseqn2002as:*
7: geneseqn2002bs:*
8: geneseqn2003as:*
9: geneseqn2003bs:*
10: geneseqn2003cs:*
11: geneseqn2003ds:*
12: geneseqn2004as:*
13: geneseqn2004bs:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	1514	100.0	1514	2	AAV65632	Aav65632 Human hea
	2	1512.4	99.9	11284	2	AAV65633	Aav65633 Plasmid c
	3	1510.8	99.8	5218	13	ACN40746	Acn40746 Tumour-as
	4	1509.8	99.7	5266	10	ADF81754	Adf81754 Leukaemia
	5	1460.2	96.4	5230	4	AAK51510	Aak51510 Human pol
	6	1455.4	96.1	4850	4	AAK52494	Aak52494 Human pol
	7	1447.4	95.6	5009	2	AAV65634	Aav65634 First Eco
	8	854.8	56.5	915	4	AAD07629	Aad07629 Human sec
	9	608.6	40.2	10331	4	AAK79915	Aak79915 Human imm
	10	608.6	40.2	10331	4	AAL03344	Aal03344 Human rep
	11	442.8	29.2	477	9	ACH35532	Ach35532 Human end
	12	441	29.1	479	9	ACH32225	Ach32225 Human end
	13	406.2	26.8	610	13	ADQ56824	Adq56824 Novel can
	14	166.4	11.0	1209	4	ABL13947	Abl13947 Drosophil
c	15	150.4	9.9	3264	4	ABL13946	Abl13946 Drosophil
	16	123.2	8.1	1545	4	AAF23894	Aaf23894 Human sec
c	17	108.6	7.2	3793	4	ABL13894	Abl13894 Drosophil
c	18	71.8	4.7	3567	4	ABL13952	Abl13952 Drosophil
	19	60	4.0	60	6	ABN44539	Abn44539 Human spl
	20	56	3.7	760	3	AAZ98210	Aaz98210 Human sig
	21	50	3.3	2000	8	ADA71938	Ada71938 Rice gene
c	22	46.2	3.1	2000	8	ADA71938	Ada71938 Rice gene
	23	42.4	2.8	1438	8	ACA31561	Aca31561 Prokaryot
c	24	41.2	2.7	114793	4	AAD08215	Aad08215 Human gen
c	25	40.6	2.7	558	5	AAH97782	Aah97782 Murine 7-
	26	40.6	2.7	2036	13	ADT44716	Adt44716 Bacterial
	27	40.4	2.7	5019	13	ADQ80232	Adq80232 Brain-spe
	28	40.4	2.7	5704	12	ADQ84111	Adq84111 Human tum
	29	40.4	2.7	5932	10	ADB75488	Adb75488 Prostate
	30	40	2.6	40	2	AAV65637	Aav65637 Primer am
	31	40	2.6	40	2	AAV65635	Aav65635 Northern
	32	40	2.6	1554	8	ADA69815	Ada69815 Rice gene
	33	39	2.6	372	8	ACA43606	Aca43606 Prokaryot
	34	39	2.6	1079	10	ADC37615	Adc37615 Human nuc
	35	39	2.6	1131	11	ABD01472	Abd01472 Pseudomon
	36	39	2.6	1209	3	AAA61286	Aaa61286 Human sec
	37	39	2.6	1209	8	ADA40219	Ada40219 Human sec
	38	39	2.6	1209	10	ADD37708	Add37708 Human sec
	39	39	2.6	1209	10	ADA56378	Ada56378 Gene enco
	40	39	2.6	1353	11	ABD01484	Abd01484 Pseudomon
	41	39	2.6	1404	10	ADC16737	Adc16737 Human mod
	42	39	2.6	1528	4	AAH15283	Aah15283 Human cDN
	43	39	2.6	1555	4	AAS44695	Aas44695 Human ful
	44	39	2.6	1566	4	AAS01567	Aas01567 Human sec
	45	39	2.6	1575	6	ADI16595	Adi16595 Human NOV

ALIGNMENTS

RESULT 1
AAV65632
ID AAV65632 standard; DNA; 1514 BP.
XX
AC AAV65632;
XX
DT 16-DEC-1998 (first entry)
XX
DE Human heart O-fucosyltransferase encoding DNA.
XX
KW O-fucosyltransferase; epidermal growth factor; EGF; glycosylation;
KW O-fucose; inhibitor; sensory neuron; retinal neuron; human; heart; ss.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT misc_feature 1. .1100
FT /*tag= b
FT /note= "this actively expressed O-fucosyltransferase
FT sequence is claimed for in claim 9"
FT CDS 1. .1098
FT /*tag= a
FT /product= "human heart O-fucosyltransferase"
XX
PN WO9833924-A1.
XX
PD 06-AUG-1998.
XX
PF 17-DEC-1997; 97WO-US023401.
XX
PR 31-JAN-1997; 97US-00792498.
PR 26-NOV-1997; 97US-00978741.
XX
PA (GETH) GENENTECH INC.
XX
PI Wang Y, Spellman MW;
XX
DR WPI; 1998-437477/37.
DR P-PSDB; AAW80571.
XX
PT Human O-fucosyl:transferase able to glycosylate epidermal growth factor
PT domains - useful for diagnosis and treatment of diseases involving
PT overexpression of the enzyme.

```
XX
PS   Claim 9; Fig 12A; 90pp; English.
XX
CC   This DNA encodes a human heart O-fucosyltransferase that can glycosylate
CC   an epidermal growth factor (EGF) domain of a polypeptide with an
CC   activated O-fucose residue. Inhibitors of O-fucosyltransferase, e.g.
CC   mutants with increased affinity for the EFG domains, are used in
CC   diagnosis and treatment of conditions associated with overexpression of O
CC   -fucosyltransferase, to promote survival of sensory (retinal) neurons.
CC   Probes based on EGF domain polypeptide are used to detect gene
CC   amplification and expression. The expression can also be determined at
CC   the protein level using antibodies specific for O-fucosyltransferase
XX
SQ   Sequence 1514 BP; 320 A; 435 C; 411 G; 348 T; 0 U; 0 Other;

  Query Match             100.0%;  Score 1514;  DB 2;  Length 1514;
  Best Local Similarity   100.0%;  Pred. No. 0;
  Matches 1514;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 ATGCCCGCGGGCTCCTGGGACCCGGCCGGTTACCTGCTCTACTGCCCCTGCATGGGGCGC 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGCCCGCGGGCTCCTGGGACCCGGCCGGTTACCTGCTCTACTGCCCCTGCATGGGGCGC 60

Qy         61 TTTGGGAACCAGGCCGATCACTTCTTGGGCTCTCTGGCATTTGCAAAGCTGCTAAACCGT 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 TTTGGGAACCAGGCCGATCACTTCTTGGGCTCTCTGGCATTTGCAAAGCTGCTAAACCGT 120

Qy        121 ACCTTGGCTGTCCCTCCTTGGATTGAGTACCAGCATCACAAGCCTCCTTTCACCAACCTC 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 ACCTTGGCTGTCCCTCCTTGGATTGAGTACCAGCATCACAAGCCTCCTTTCACCAACCTC 180

Qy        181 CATGTGTCCTACCAGAAGTACTTCAAGCTGGAGCCCCTCCAGGCTTACCATCGGGTCATC 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 CATGTGTCCTACCAGAAGTACTTCAAGCTGGAGCCCCTCCAGGCTTACCATCGGGTCATC 240

Qy        241 AGCTTGGAGGATTTCATGGAGAAGCTGGCACCCACCCACTGGCCCCCTGAGAAGCGGGTG 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 AGCTTGGAGGATTTCATGGAGAAGCTGGCACCCACCCACTGGCCCCCTGAGAAGCGGGTG 300

Qy        301 GCATACTGCTTTGAGGTGGCAGCCCAGCGAAGCCCAGATAAGAAGACGTGCCCCATGAAG 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 GCATACTGCTTTGAGGTGGCAGCCCAGCGAAGCCCAGATAAGAAGACGTGCCCCATGAAG 360

Qy        361 GAAGGAAACCCCTTTGGCCCATTCTGGGATCAGTTTCATGTGAGTTTCAACAAGTCGGAG 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 GAAGGAAACCCCTTTGGCCCATTCTGGGATCAGTTTCATGTGAGTTTCAACAAGTCGGAG 420

Qy        421 CTTTTTACAGGCATTTCCTTCAGTGCTTCCTACAGAGAACAATGGAGCCAGAGATTTTCT 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 CTTTTTACAGGCATTTCCTTCAGTGCTTCCTACAGAGAACAATGGAGCCAGAGATTTTCT 480

Qy        481 CCAAAGGAACATCCGGTGCTTGCCCTGCCAGGAGCCCCAGCCCAGTTCCCCGTCCTAGAA 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 CCAAAGGAACATCCGGTGCTTGCCCTGCCAGGAGCCCCAGCCCAGTTCCCCGTCCTAGAA 540

Qy        541 GAACACAGGCCACTACAGAAGTACATGGTATGGTCAGACGAAATGGTGAAGACGGGAGAG 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 GAACACAGGCCACTACAGAAGTACATGGTATGGTCAGACGAAATGGTGAAGACGGGAGAG 600

Qy        601 GCCCAGATTCATGCCCACCTTGTCCGGCCCTATGTGGGCATTCATCTGCGCATTGGCTCT 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 GCCCAGATTCATGCCCACCTTGTCCGGCCCTATGTGGGCATTCATCTGCGCATTGGCTCT 660

Qy        661 GACTGGAAGAACGCCTGTGCCATGCTGAAGGACGGGACTGCAGGCTCGCACTTCATGGCC 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 GACTGGAAGAACGCCTGTGCCATGCTGAAGGACGGGACTGCAGGCTCGCACTTCATGGCC 720

Qy        721 TCTCCGCAGTGTGTGGGCTACAGCCGCAGCACAGCGGCCCCCCTCACGATGACTATGTGC 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 TCTCCGCAGTGTGTGGGCTACAGCCGCAGCACAGCGGCCCCCCTCACGATGACTATGTGC 780

Qy        781 CTGCCTGACCTGAAGGAGATCCAGAGGGCTGTGAAGCTCTGGGTGAGGTCGCTGGATGCC 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db        781 CTGCCTGACCTGAAGGAGATCCAGAGGGCTGTGAAGCTCTGGGTGAGGTCGCTGGATGCC 840

Qy        841 CAGTCGGTCTACGTTGCTACTGATTCCGAGAGTTATGTGCCTGAGCTCCAACAGCTCTTC 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 CAGTCGGTCTACGTTGCTACTGATTCCGAGAGTTATGTGCCTGAGCTCCAACAGCTCTTC 900

Qy        901 AAAGGGAAGGTGAAGGTGGTGAGCCTGAAGCCTGAGGTGGCCCAGGTCGACCTGTACATC 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 AAAGGGAAGGTGAAGGTGGTGAGCCTGAAGCCTGAGGTGGCCCAGGTCGACCTGTACATC 960

Qy        961 CTCGGCCAAGCCGACCACTTTATTGGCAACTGTGTCTCCTCCTTCACTGCCTTTGTGAAG 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 CTCGGCCAAGCCGACCACTTTATTGGCAACTGTGTCTCCTCCTTCACTGCCTTTGTGAAG 1020

Qy       1021 CGGGAGCGGGACCTCCAGGGGAGGCCGTCTTCTTTCTTCGGCATGGACAGGCCCCCTAAG 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1021 CGGGAGCGGGACCTCCAGGGGAGGCCGTCTTCTTTCTTCGGCATGGACAGGCCCCCTAAG 1080

Qy       1081 CTGCGGGACGAGTTCTGATTCTGGCCGGAGCACCAGACCCTCTGATCCTGGAGGGACCAG 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1081 CTGCGGGACGAGTTCTGATTCTGGCCGGAGCACCAGACCCTCTGATCCTGGAGGGACCAG 1140

Qy       1141 AGTCTGAGCTGGTCCTTCCAGCCAGGCCTGGCAGCCAGAGGTGCTCCGGGATTGCAAACT 1200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1141 AGTCTGAGCTGGTCCTTCCAGCCAGGCCTGGCAGCCAGAGGTGCTCCGGGATTGCAAACT 1200

Qy       1201 CCTCTTCTCACCTGCCAAAGATGGAGAAGAGTGCCAGGGACCCCTCAAGGAGGGAGACGC 1260
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1201 CCTCTTCTCACCTGCCAAAGATGGAGAAGAGTGCCAGGGACCCCTCAAGGAGGGAGACGC 1260

Qy       1261 TCCATATCCCAGGGCATAGGACTTGCAGGTTCCTAGGAGCAGGAGCATCTCCCATCGCAC 1320
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1261 TCCATATCCCAGGGCATAGGACTTGCAGGTTCCTAGGAGCAGGAGCATCTCCCATCGCAC 1320

Qy       1321 GTGCTTTCTGCTCTTCTGGGAATTTCTCACACTGGCAAAGCAGTCCAGCCTCCGTCTTCT 1380
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1321 GTGCTTTCTGCTCTTCTGGGAATTTCTCACACTGGCAAAGCAGTCCAGCCTCCGTCTTCT 1380

Qy       1381 GGTCCACTCTGCTCTGAGCAGCCTGGGATGCTGAACTCTTCAGAGAGATTTTTTTATAGA 1440
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1381 GGTCCACTCTGCTCTGAGCAGCCTGGGATGCTGAACTCTTCAGAGAGATTTTTTTATAGA 1440

Qy       1441 GAGATTTCTATAATTTTGATACAAGGTCATGACTATCCTAGAACTCTCTGTGGTTTTTGA 1500
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1441 GAGATTTCTATAATTTTGATACAAGGTCATGACTATCCTAGAACTCTCTGTGGTTTTTGA 1500

Qy       1501 AAATCATTGAATTC 1514
              ||||||||||||||
Db       1501 AAATCATTGAATTC 1514
```

```
RESULT 2
AAV65633
ID   AAV65633 standard; DNA; 11284 BP.
XX
AC   AAV65633;
XX
DT   16-DEC-1998  (first entry)
XX
DE   Plasmid construct for expression of human O-fucosyltransferase.
XX
KW   O-fucosyltransferase; epidermal growth factor; EGF; glycosylation;
KW   O-fucose; inhibitor; sensory neuron; retinal neuron; human; heart; ss.
XX
OS   Homo sapiens.
OS   Synthetic.
XX
FH   Key             Location/Qualifiers
FT   CDS             4140. .5333
FT                   /*tag= a
FT                   /note= "insert coding for human O-fucosyltransferase."
FT   misc_feature    4218. .4235
FT                   /*tag= b
```

```
FT                   /note= "polyhistidine tag"
XX
PN   WO9833924-A1.
XX
PD   06-AUG-1998.
XX
PF   17-DEC-1997;   97WO-US023401.
XX
PR   31-JAN-1997;   97US-00792498.
PR   26-NOV-1997;   97US-00978741.
XX
PA   (GETH ) GENENTECH INC.
XX
PI   Wang Y,  Spellman MW;
XX
DR   WPI; 1998-437477/37.
DR   P-PSDB; AAW80573.
XX
PT   Human O-fucosyl:transferase able to glycosylate epidermal growth factor
PT   domains - useful for diagnosis and treatment of diseases involving
PT   overexpression of the enzyme.
XX
PS   Example; Page 49-57; 90pp; English.
XX
CC   This represents the nucleotide sequence of the plasmid construct used for
CC   the expression of human heart O-fucosyltransferase. The human O-
CC   fucosyltransferase can glycosylate an epidermal growth factor (EGF)
CC   domain of a polypeptide with an activated O-fucose residue. Inhibitors of
CC   O-fucosyltransferase, e.g. mutants with increased affinity for the EFG
CC   domains, are used in diagnosis and treatment of conditions associated
CC   with overexpression of O-fucosyltransferase, to promote survival of
CC   sensory (retinal) neurons. Probes based on EGF domain polypeptide are
CC   used to detect gene amplification and expression. The expression can also
CC   be determined at the protein level using antibodies specific for O-
CC   fucosyltransferase
XX
SQ   Sequence 11284 BP; 2947 A; 2602 C; 2624 G; 3111 T; 0 U; 0 Other;

  Query Match             99.9%;  Score 1512.4;  DB 2;  Length 11284;
  Best Local Similarity   99.9%;  Pred. No. 0;
  Matches 1513;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 ATGCCCGCGGGGCTCCTGGGACCCGGCCGGTTACCTGCTCTACTGCCCCTGCATGGGGCGC 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4236 ATGCCCGCGGGGCTCCTGGGACCCGGCCGGTTACCTGCTCTACTGCCCCTGCATGGGGCGC 4295

Qy         61 TTTGGGAACCAGGCCGATCACTTCTTGGGCTCTCTGGCATTTGCAAAGCTGCTAAACCGT 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4296 TTTGGGAACCAGGCCGATCACTTCTTGGGCTCTCTGGCATTTGCAAAGCTGCTAAACCGT 4355

Qy        121 ACCTTGGCTGTCCCTCCTTGGATTGAGTACCAGCATCACAAGCCTCCTTTCACCAACCTC 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4356 ACCTTGGCTGTCCCTCCTTGGATTGAGTACCAGCATCACAAGCCTCCTTTCACCAACCTC 4415

Qy        181 CATGTGTCCTACCAGAAGTACTTCAAGCTGGAGCCCCTCCAGGCTTACCATCGGGTCATC 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4416 CATGTGTCCTACCAGAAGTACTTCAAGCTGGAGCCCCTCCAGGCTTACCATCGGGTCATC 4475

Qy        241 AGCTTGGAGGATTTCATGGAGAAGCTGGCACCCACCCACTGGCCCCCTGAGAAGCGGGTG 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4476 AGCTTGGAGGATTTCATGGAGAAGCTGGCACCCACCCACTGGCCCCCTGAGAAGCGGGTG 4535

Qy        301 GCATACTGCTTTGAGGTGGCAGCCCAGCGAAGCCCAGATAAGAAGACGTGCCCCATGAAG 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4536 GCATACTGCTTTGAGGTGGCAGCCCAGCGAAGCCCAGATAAGAAGACGTGCCCCATGAAG 4595

Qy        361 GAAGGAAACCCCTTTGGCCCATTCTGGGATCAGTTTCATGTGAGTTTCAACAAGTCGGAG 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4596 GAAGGAAACCCCTTTGGCCCATTCTGGGATCAGTTTCATGTGAGTTTCAACAAGTCGGAG 4655

Qy        421 CTTTTTACAGGCATTTCCTTCAGTGCTTCCTACAGAGAACAATGGAGCCAGAGATTTTCT 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4656 CTTTTTACAGGCATTTCCTTCAGTGCTTCCTACAGAGAACAATGGAGCCAGAGATTTTCT 4715

Qy        481 CCAAAGGAACATCCGGTGCTTGCCCTGCCAGGAGCCCCAGCCCAGTTCCCCGTCCTAGAA 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4716 CCAAAGGAACATCCGGTGCTTGCCCTGCCAGGAGCCCCAGCCCAGTTCCCCGTCCTAGAG 4775

Qy        541 GAACACAGGCCACTACAGAAGTACATGGTATGGTCAGACGAAATGGTGAAGACGGGAGAG 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4776 GAACACAGGCCACTACAGAAGTACATGGTATGGTCAGACGAAATGGTGAAGACGGGAGAG 4835

Qy        601 GCCCAGATTCATGCCCACCTTGTCCGGCCCTATGTGGGCATTCATCTGCGCATTGGCTCT 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4836 GCCCAGATTCATGCCCACCTTGTCCGGCCCTATGTGGGCATTCATCTGCGCATTGGCTCT 4895

Qy        661 GACTGGAAGAACGCCTGTGCCATGCTGAAGGACGGGACTGCAGGCTCGCACTTCATGGCC 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4896 GACTGGAAGAACGCCTGTGCCATGCTGAAGGACGGGACTGCAGGCTCGCACTTCATGGCC 4955

Qy        721 TCTCCGCAGTGTGTGGGCTACAGCCGCAGCACAGCGGCCCCCCTCACGATGACTATGTGC 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4956 TCTCCGCAGTGTGTGGGCTACAGCCGCAGCACAGCGGCCCCCCTCACGATGACTATGTGC 5015

Qy        781 CTGCCTGACCTGAAGGAGATCCAGAGGGCTGTGAAGCTCTGGGTGAGGTCGCTGGATGCC 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5016 CTGCCTGACCTGAAGGAGATCCAGAGGGCTGTGAAGCTCTGGGTGAGGTCGCTGGATGCC 5075

Qy        841 CAGTCGGTCTACGTTGCTACTGATTCCGAGAGTTATGTGCCTGAGCTCCAACAGCTCTTC 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5076 CAGTCGGTCTACGTTGCTACTGATTCCGAGAGTTATGTGCCTGAGCTCCAACAGCTCTTC 5135

Qy        901 AAAGGGAAGGTGAAGGTGGTGAGCCTGAAGCCTGAGGTGGCCCAGGTCGACCTGTACATC 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5136 AAAGGGAAGGTGAAGGTGGTGAGCCTGAAGCCTGAGGTGGCCCAGGTCGACCTGTACATC 5195

Qy        961 CTCGGCCAAGCCGACCACTTTATTGGCAACTGTGTCTCCTCCTTCACTGCCTTTGTGAAG 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5196 CTCGGCCAAGCCGACCACTTTATTGGCAACTGTGTCTCCTCCTTCACTGCCTTTGTGAAG 5255

Qy       1021 CGGGAGCGGGACCTCCAGGGGAGGCCGTCTTCTTTCTTCGGCATGGACAGGCCCCCTAAG 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5256 CGGGAGCGGGACCTCCAGGGGAGGCCGTCTTCTTTCTTCGGCATGGACAGGCCCCCTAAG 5315

Qy       1081 CTGCGGGACGAGTTCTGATTCTGGCCGGAGCACCAGACCCTCTGATCCTGGAGGGACCAG 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5316 CTGCGGGACGAGTTCTGATTCTGGCCGGAGCACCAGACCCTCTGATCCTGGAGGGACCAG 5375

Qy       1141 AGTCTGAGCTGGTCCTTCCAGCCAGGCCTGGCAGCCAGAGGTGCTCCGGGATTGCAAACT 1200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5376 AGTCTGAGCTGGTCCTTCCAGCCAGGCCTGGCAGCCAGAGGTGCTCCGGGATTGCAAACT 5435

Qy       1201 CCTCTTCTCACCTGCCAAAGATGGAGAAGAGTGCCAGGGACCCCTCAAGGAGGGAGACGC 1260
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5436 CCTCTTCTCACCTGCCAAAGATGGAGAAGAGTGCCAGGGACCCCTCAAGGAGGGAGACGC 5495

Qy       1261 TCCATATCCCAGGGCATAGGACTTGCAGGTTCCTAGGAGCAGGAGCATCTCCCATCGCAC 1320
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5496 TCCATATCCCAGGGCATAGGACTTGCAGGTTCCTAGGAGCAGGAGCATCTCCCATCGCAC 5555

Qy       1321 GTGCTTTCTGCTCTTCTGGGAATTTCTCACACTGGCAAAGCAGTCCAGCCTCCGTCTTCT 1380
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5556 GTGCTTTCTGCTCTTCTGGGAATTTCTCACACTGGCAAAGCAGTCCAGCCTCCGTCTTCT 5615

Qy       1381 GGTCCACTCTGCTCTGAGCAGCCTGGGATGCTGAACTCTTCAGAGAGATTTTTTTATAGA 1440
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5616 GGTCCACTCTGCTCTGAGCAGCCTGGGATGCTGAACTCTTCAGAGAGATTTTTTTATAGA 5675

Qy       1441 GAGATTTCTATAATTTTGATACAAGGTCATGACTATCCTAGAACTCTCTGTGGTTTTTGA 1500
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5676 GAGATTTCTATAATTTTGATACAAGGTCATGACTATCCTAGAACTCTCTGTGGTTTTTGA 5735

Qy       1501 AAATCATTGAATTC 1514
              ||||||||||||||
Db       5736 AAATCATTGAATTC 5749
```

RESULT 3
ACN40746
ID ACN40746 standard; cDNA; 5218 BP.
XX
AC ACN40746;
XX
DT 18-NOV-2004 (first entry)
XX
DE Tumour-associated antigenic target (TAT) cDNA DNA326813, SEQ ID NO:5725.
XX
KW Tumour-associated antigenic target; TAT; human; overexpression; cancer;
KW tumour; diagnosis; cell proliferative disorder; breast cancer;
KW colorectal cancer; lung cancer; ovarian cancer; liver cancer;
KW central nervous system cancer; bladder cancer; pancreatic cancer;
KW cervical cancer; melanoma; leukaemia; hybridisation probe;
KW chromosome identification; chromosome mapping; gene mapping;
KW gene therapy; cytostatic; gene; ss.
XX
OS Homo sapiens.
XX
PN WO2004030615-A2.
XX
PD 15-APR-2004.
XX
PF 29-SEP-2003; 2003WO-US028547.
XX
PR 02-OCT-2002; 2002US-0414971P.
XX
PA (GETH) GENENTECH INC.
XX
PI Wu TD, Zhang Z, Zhou Y;
XX
DR WPI; 2004-347921/32.
DR P-PSDB; ABM82223.
XX
PT New tumor-associated antigenic target polypeptides and nucleic acids,
PT useful in preparing a medicament for treating or detecting a
PT proliferative disorder, e.g. breast, lung, colorectal, ovarian or
PT prostate cancer or tumor.
XX
PS Claim 1; SEQ ID NO 5725; 7273pp; English.
XX
CC The invention relates to human tumour-associated antigenic target (TAT)
CC polypeptides, and their related nucleic acids. The TAT polypeptides are
CC overexpressed in cancer tissues compared to normal tissues, and may thus
CC serve as effective targets for the diagnosis and treatment of cancer in
CC mammals. The invention also relates to nucleic acid and polypeptide
CC sequences at least 80% identical to the TAT nucleic acids and
CC polypeptides; expression vectors and host cells comprising a TAT nucleic
CC acid; an antibody specific for a TAT polypeptide; a peptide or organic
CC molecule which binds to a TAT polypeptide; fusion proteins comprising a
CC TAT polypeptide; and methods and compositions for the treatment or
CC diagnosis of cancer in mammals. TAT polypeptides, nucleic acids,
CC antibodies, antagonists, binding molecules and compositions are useful
CC for diagnosing or treating a cell proliferative disorder associated with
CC increased TAT expression, particularly cancers such as breast cancer,
CC colorectal cancer, lung cancer, ovarian cancer, liver cancer, bladder
CC cancer, pancreatic cancer, cervical cancer, cancers of the central
CC nervous system, melanoma and leukaemia. TAT nucleic acids may further be
CC used as hybridisation probes, in chromosome and gene mapping, in
CC chromosome identification and in gene therapy. The present sequence
CC represents a TAT nucleic acid of the invention
XX
SQ Sequence 5218 BP; 1255 A; 1274 C; 1334 G; 1355 T; 0 U; 0 Other;

Query Match 99.8%; Score 1510.8; DB 13; Length 5218;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 1512; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

```
Qy          1 ATGCCCGCGGGCTCCTGGGACCCGGCCGGTTACCTGCTCTACTGCCCCTGCATGGGGCGC 60
              ||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        119 ATGCCTGCGGGCTCCTGGGACCCGGCCGGTTACCTGCTCTACTGCCCCTGCATGGGGCGC 178

Qy         61 TTTGGGAACCAGGCCGATCACTTCTTGGGCTCTCTGGCATTTGCAAAGCTGCTAAACCGT 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        179 TTTGGGAACCAGGCCGATCACTTCTTGGGCTCTCTGGCATTTGCAAAGCTGCTAAACCGT 238

Qy        121 ACCTTGGCTGTCCCTCCTTGGATTGAGTACCAGCATCACAAGCCTCCTTTCACCAACCTC 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        239 ACCTTGGCTGTCCCTCCTTGGATTGAGTACCAGCATCACAAGCCTCCTTTCACCAACCTC 298

Qy        181 CATGTGTCCTACCAGAAGTACTTCAAGCTGGAGCCCCTCCAGGCTTACCATCGGGTCATC 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        299 CATGTGTCCTACCAGAAGTACTTCAAGCTGGAGCCCCTCCAGGCTTACCATCGGGTCATC 358

Qy        241 AGCTTGGAGGATTTCATGGAGAAGCTGGCACCCACCCACTGGCCCCCTGAGAAGCGGGTG 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        359 AGCTTGGAGGATTTCATGGAGAAGCTGGCACCCACCCACTGGCCCCCTGAGAAGCGGGTG 418

Qy        301 GCATACTGCTTTGAGGTGGCAGCCCAGCGAAGCCCAGATAAGAAGACGTGCCCCATGAAG 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        419 GCATACTGCTTTGAGGTGGCAGCCCAGCGAAGCCCAGATAAGAAGACGTGCCCCATGAAG 478

Qy        361 GAAGGAAACCCCTTTGGCCCATTCTGGGATCAGTTTCATGTGAGTTTCAACAAGTCGGAG 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        479 GAAGGAAACCCCTTTGGCCCATTCTGGGATCAGTTTCATGTGAGTTTCAACAAGTCGGAG 538

Qy        421 CTTTTTACAGGCATTTCCTTCAGTGCTTCCTACAGAGAACAATGGAGCCAGAGATTTTCT 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        539 CTTTTTACAGGCATTTCCTTCAGTGCTTCCTACAGAGAACAATGGAGCCAGAGATTTTCT 598

Qy        481 CCAAAGGAACATCCGGTGCTTGCCCTGCCAGGAGCCCCAGCCCAGTTCCCCGTCCTAGAA 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        599 CCAAAGGAACATCCGGTGCTTGCCCTGCCAGGAGCCCCAGCCCAGTTCCCCGTCCTAGAG 658

Qy        541 GAACACAGGCCACTACAGAAGTACATGGTATGGTCAGACGAAATGGTGAAGACGGGAGAG 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        659 GAACACAGGCCACTACAGAAGTACATGGTATGGTCAGACGAAATGGTGAAGACGGGAGAG 718

Qy        601 GCCCAGATTCATGCCCACCTTGTCCGGCCCTATGTGGGCATTCATCTGCGCATTGGCTCT 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        719 GCCCAGATTCATGCCCACCTTGTCCGGCCCTATGTGGGCATTCATCTGCGCATTGGCTCT 778

Qy        661 GACTGGAAGAACGCCTGTGCCATGCTGAAGGACGGGACTGCAGGCTCGCACTTCATGGCC 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        779 GACTGGAAGAACGCCTGTGCCATGCTGAAGGACGGGACTGCAGGCTCGCACTTCATGGCC 838

Qy        721 TCTCCGCAGTGTGTGGGCTACAGCCGCAGCACAGCGGCCCCCCTCACGATGACTATGTGC 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        839 TCTCCGCAGTGTGTGGGCTACAGCCGCAGCACAGCGGCCCCCCTCACGATGACTATGTGC 898

Qy        781 CTGCCTGACCTGAAGGAGATCCAGAGGGCTGTGAAGCTCTGGGTGAGGTCGCTGGATGCC 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        899 CTGCCTGACCTGAAGGAGATCCAGAGGGCTGTGAAGCTCTGGGTGAGGTCGCTGGATGCC 958

Qy        841 CAGTCGGTCTACGTTGCTACTGATTCCGAGAGTTATGTGCCTGAGCTCCAACAGCTCTTC 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        959 CAGTCGGTCTACGTTGCTACTGATTCCGAGAGTTATGTGCCTGAGCTCCAACAGCTCTTC 1018

Qy        901 AAAGGGAAGGTGAAGGTGGTGAGCCTGAAGCCTGAGGTGGCCCAGGTCGACCTGTACATC 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1019 AAAGGGAAGGTGAAGGTGGTGAGCCTGAAGCCTGAGGTGGCCCAGGTCGACCTGTACATC 1078

Qy        961 CTCGGCCAAGCCGACCACTTTATTGGCAACTGTGTCTCCTCCTTCACTGCCTTTGTGAAG 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1079 CTCGGCCAAGCCGACCACTTTATTGGCAACTGTGTCTCCTCCTTCACTGCCTTTGTGAAG 1138

Qy       1021 CGGGAGCGGGACCTCCAGGGGAGGCCGTCTTCTTTCTTCGGCATGGACAGGCCCCCTAAG 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1139 CGGGAGCGGGACCTCCAGGGGAGGCCGTCTTCTTTCTTCGGCATGGACAGGCCCCCTAAG 1198

Qy       1081 CTGCGGGACGAGTTCTGATTCTGGCCGGAGCACCAGACCCTCTGATCCTGGAGGGACCAG 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1199 CTGCGGGACGAGTTCTGATTCTGGCCGGAGCACCAGACCCTCTGATCCTGGAGGGACCAG 1258
```

```
Qy       1141 AGTCTGAGCTGGTCCTTCCAGCCAGGCCTGGCAGCCAGAGGTGCTCCGGGATTGCAAACT 1200
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1259 AGTCTGAGCTGGTCCTTCCAGCCAGGCCTGGCAGCCAGAGGTGCTCCGGGATTGCAAACT 1318

Qy       1201 CCTCTTCTCACCTGCCAAAGATGGAGAAGAGTGCCAGGGACCCCTCAAGGAGGGAGACGC 1260
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1319 CCTCTTCTCACCTGCCAAAGATGGAGAAGAGTGCCAGGGACCCCTCAAGGAGGGAGACGC 1378

Qy       1261 TCCATATCCCAGGGCATAGGACTTGCAGGTTCCTAGGAGCAGGAGCATCTCCCATCGCAC 1320
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1379 TCCATATCCCAGGGCATAGGACTTGCAGGTTCCTAGGAGCAGGAGCATCTCCCATCGCAC 1438

Qy       1321 GTGCTTTCTGCTCTTCTGGGAATTTCTCACACTGGCAAAGCAGTCCAGCCTCCGTCTTCT 1380
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1439 GTGCTTTCTGCTCTTCTGGGAATTTCTCACACTGGCAAAGCAGTCCAGCCTCCGTCTTCT 1498

Qy       1381 GGTCCACTCTGCTCTGAGCAGCCTGGGATGCTGAACTCTTCAGAGAGATTTTTTTATAGA 1440
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1499 GGTCCACTCTGCTCTGAGCAGCCTGGGATGCTGAACTCTTCAGAGAGATTTTTTTATAGA 1558

Qy       1441 GAGATTTCTATAATTTTGATACAAGGTCATGACTATCCTAGAACTCTCTGTGGTTTTTGA 1500
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1559 GAGATTTCTATAATTTTGATACAAGGTCATGACTATCCTAGAACTCTCTGTGGTTTTTGA 1618

Qy       1501 AAATCATTGAATTC 1514
             ||||||||||||||
Db       1619 AAATCATTGAATTC 1632
```

RESULT 4
ADF81754
ID ADF81754 standard; DNA; 5266 BP.
XX
AC ADF81754;
XX
DT 26-FEB-2004 (first entry)
XX
DE Leukaemia-related DNA sequence #2310.
XX
KW Cytostatic; Gene therapy; leukaemia; ss.
XX
OS Unidentified.
XX
PN WO2003039443-A2.
XX
PD 15-MAY-2003.
XX
PF 04-NOV-2002; 2002WO-EP012303.
XX
PR 05-NOV-2001; 2001EP-00126244.
PR 30-APR-2002; 2002EP-00009758.
XX
PA (DEKR-) DEUT KREBSFORSCHUNGSZENTRUM.
PA (UYLU-) UNIV LUDWIG MAXIMILIANS.
PA (HAFE/) HAFERLACH T.
PA (SCHO/) SCHOCH C.
PA (KERN/) KERN W.
XX
PI Haferlach T, Schoch C, Kern W, Kohlmann A, Schnittger S, Dugas M;
PI Eils R, Brors B, Mergenthaler S;
XX
DR WPI; 2003-505037/47.
XX
PT Determining the subtype of leukemia cells and whether a patient sample
PT contains leukemia cells or other cells, useful for treating leukemia,
PT comprises determining the expression profile of a group of markers in a
PT patient sample.
XX
PS Disclosure; SEQ ID NO 2310; 2938pp; English.
XX
CC The present invention relates to a method (M1) for determining the
CC subtype of leukaemia cells and whether a patient sample contains
CC leukaemia cells. The method comprises determining the expression profile
CC of a group of markers in a patient sample. The method is useful for
CC determining the presence of leukaemia cells, its types or subtypes, and
CC for the preparation of a medicament for treating leukaemia.
XX
SQ Sequence 5266 BP; 1280 A; 1278 C; 1331 G; 1346 T; 0 U; 31 Other;

```
  Query Match            99.7%;  Score 1509.8;  DB 10;  Length 5266;
  Best Local Similarity  99.8%;  Pred. No. 0;
  Matches 1511;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy          1 ATGCCCGCGGGCTCCTGGGACCCGGCCGGTTACCTGCTCTACTGCCCCTGCATGGGGCGC 60
              ||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        132 ATGCCTGCGGGCTCCTGGGACCCGGCCGGTTACCTGCTCTACTGCCCCTGCATGGGGCGC 191

Qy         61 TTTGGGAACCAGGCCGATCACTTCTTGGGCTCTCTGGCATTTGCAAAGCTGCTAAACCGT 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        192 TTTGGGAACCAGGCCGATCACTTCTTGGGCTCTCTGGCATTTGCAAAGCTGCTAAACCGT 251

Qy        121 ACCTTGGCTGTCCCTCCTTGGATTGAGTACCAGCATCACAAGCCTCCTTTCACCAACCTC 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        252 ACCTTGGCTGTCCCTCCTTGGATTGAGTACCAGCATCACAAGCCTCCTTTCACCAACCTC 311

Qy        181 CATGTGTCCTACCAGAAGTACTTCAAGCTGGAGCCCCTCCAGGCTTACCATCGGGTCATC 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        312 CATGTGTCCTACCAGAAGTACTTCAAGCTGGAGCCCCTCCAGGCTTACCATCGGGTCATC 371

Qy        241 AGCTTGGAGGATTTCATGGAGAAGCTGGCACCCACCCACTGGCCCCCTGAGAAGCGGGTG 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        372 AGCTTGGAGGATTTCATGGAGAAGCTGGCACCCACCCACTGGCCCCCTGAGAAGCGGGTG 431

Qy        301 GCATACTGCTTTGAGGTGGCAGCCCAGCGAAGCCCAGATAAGAAGACGTGCCCCATGAAG 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        432 GCATACTGCTTTGAGGTGGCAGCCCAGCGAAGCCCAGATAAGAAGACGTGCCCCATGAAG 491

Qy        361 GAAGGAAACCCCTTTGGCCCATTCTGGGATCAGTTTCATGTGAGTTTCAACAAGTCGGAG 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        492 GAAGGAAACCCCTTTGGCCCATTCTGGGATCAGTTTCATGTGAGTTTCAACAAGTCGGAG 551

Qy        421 CTTTTTACAGGCATTTCCTTCAGTGCTTCCTACAGAGAACAATGGAGCCAGAGATTTTCT 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        552 CTTTTTACAGGCATTTCCTTCAGTGCTTCCTACAGAGAACAATGGAGCCAGAGATTTTCT 611

Qy        481 CCAAAGGAACATCCGGTGCTTGCCCTGCCAGGAGCCCCAGCCCAGTTCCCCGTCCTAGAA 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        612 CCAAAGGAACATCCGGTGCTTGCCCTGCCAGGAGCCCCAGCCCAGTTCCCCGTCCTAGAG 671

Qy        541 GAACACAGGCCACTACAGAAGTACATGGTATGGTCAGACGAAATGGTGAAGACGGGAGAG 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        672 GAACACAGGCCACTACAGAAGTACATGGTATGGTCAGACGAAATGGTGAAGACGGGAGAG 731

Qy        601 GCCCAGATTCATGCCCACCTTGTCCGGCCCTATGTGGGCATTCATCTGCGCATTGGCTCT 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        732 GCCCAGATTCATGCCCACCTTGTCCGGCCCTATGTGGGCATTCATCTGCGCATTGGCTCT 791

Qy        661 GACTGGAAGAACGCCTGTGCCATGCTGAAGGACGGGACTGCAGGCTCGCACTTCATGGCC 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        792 GACTGGAAGAACGCCTGTGCCATGCTGAAGGACGGGACTGCAGGCTCGCACTTCATGGCC 851

Qy        721 TCTCCGCAGTGTGTGGGCTACAGCCGCAGCACAGCGGCCCCCCTCACGATGACTATGTGC 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        852 TCTCCGCAGTGTGTGGGCTACAGCCGCAGCACAGCGGCCCCCCTCACGATGACTATGTGC 911

Qy        781 CTGCCTGACCTGAAGGAGATCCAGAGGGCTGTGAAGCTCTGGGTGAGGTCGCTGGATGCC 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        912 CTGCCTGACCTGAAGGAGATCCAGAGGGCTGTGAAGCTCTGGGTGAGGTCGCTGGATGCC 971

Qy        841 CAGTCGGTCTACGTTGCTACTGATTCCGAGAGTTATGTGCCTGAGCTCCAACAGCTCTTC 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        972 CAGTCGGTCTACGTTGCTACTGATTCCGAGAGTTATGTGCCTGAGCTCCAACAGCTCTTC 1031

Qy        901 AAAGGGAAGGTGAAGGTGGTGAGCCTGAAGCCTGAGGTGGCCCAGGTCGACCTGTACATC 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1032 AAAGGGAAGGTGAAGGTGGTGAGCCTGAAGCCTGAGGTGGCCCAGGTCGACCTGTACATC 1091
```

```
Qy        961 CTCGGCCAAGCCGACCACTTTATTGGCAACTGTGTCTCCTCCTTCACTGCCTTTGTGAAG 1020
              ||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1092 CTCGGNCAAGCCGACCACTTTATTGGCAACTGTGTCTCCTCCTTCACTGCCTTTGTGAAG 1151

Qy       1021 CGGGAGCGGGACCTCCAGGGGAGGCCGTCTTCTTTCTTCGGCATGGACAGGCCCCCTAAG 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1152 CGGGAGCGGGACCTCCAGGGGAGGCCGTCTTCTTTCTTCGGCATGGACAGGCCCCCTAAG 1211

Qy       1081 CTGCGGGACGAGTTCTGATTCTGGCCGGAGCACCAGACCCTCTGATCCTGGAGGGACCAG 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1212 CTGCGGGACGAGTTCTGATTCTGGCCGGAGCACCAGACCCTCTGATCCTGGAGGGACCAG 1271

Qy       1141 AGTCTGAGCTGGTCCTTCCAGCCAGGCCTGGCAGCCAGAGGTGCTCCGGGATTGCAAACT 1200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1272 AGTCTGAGCTGGTCCTTCCAGCCAGGCCTGGCAGCCAGAGGTGCTCCGGGATTGCAAACT 1331

Qy       1201 CCTCTTCTCACCTGCCAAAGATGGAGAAGAGTGCCAGGGACCCCTCAAGGAGGGAGACGC 1260
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1332 CCTCTTCTCACCTGCCAAAGATGGAGAAGAGTGCCAGGGACCCCTCAAGGAGGGAGACGC 1391

Qy       1261 TCCATATCCCAGGGCATAGGACTTGCAGGTTCCTAGGAGCAGGAGCATCTCCCATCGCAC 1320
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1392 TCCATATCCCAGGGCATAGGACTTGCAGGTTCCTAGGAGCAGGAGCATCTCCCATCGCAC 1451

Qy       1321 GTGCTTTCTGCTCTTCTGGGAATTTCTCACACTGGCAAAGCAGTCCAGCCTCCGTCTTCT 1380
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1452 GTGCTTTCTGCTCTTCTGGGAATTTCTCACACTGGCAAAGCAGTCCAGCCTCCGTCTTCT 1511

Qy       1381 GGTCCACTCTGCTCTGAGCAGCCTGGGATGCTGAACTCTTCAGAGAGATTTTTTTATAGA 1440
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1512 GGTCCACTCTGCTCTGAGCAGCCTGGGATGCTGAACTCTTCAGAGAGATTTTTTTATAGA 1571

Qy       1441 GAGATTTCTATAATTTTGATACAAGGTCATGACTATCCTAGAACTCTCTGTGGTTTTTGA 1500
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1572 GAGATTTCTATAATTTTGATACAAGGTCATGACTATCCTAGAACTCTCTGTGGTTTTTGA 1631

Qy       1501 AAATCATTGAATTC 1514
              ||||||||||||||
Db       1632 AAATCATTGAATTC 1645
```

RESULT 5
AAK51510
ID AAK51510 standard; cDNA; 5230 BP.
XX
AC AAK51510;
XX
DT 06-NOV-2001 (first entry)
XX
DE Human polynucleotide SEQ ID NO 55.
XX
KW Human; cytokine; cell proliferation; cell differentiation; gene therapy;
KW vaccine; peptide therapy; stem cell growth factor; haematopoiesis;
KW tissue growth factor; immunomodulatory; cancer; leukaemia;
KW nervous system disorder; arthritis; inflammation; ss.
XX
OS Homo sapiens.
XX
PN WO200157190-A2.
XX
PD 09-AUG-2001.
XX
PF 05-FEB-2001; 2001WO-US004098.
XX
PR 03-FEB-2000; 2000US-00496914.
PR 27-APR-2000; 2000US-00560875.
PR 20-JUN-2000; 2000US-00598075.
PR 19-JUL-2000; 2000US-00620325.
PR 01-SEP-2000; 2000US-00654936.
PR 15-SEP-2000; 2000US-00663561.
PR 20-OCT-2000; 2000US-00693325.
PR 30-NOV-2000; 2000US-00728422.
XX
PA (HYSE-) HYSEQ INC.
XX
PI Tang YT, Liu C, Drmanac RT, Asundi V, Zhou P, Xu C, Cao Y;
PI Ma Y, Zhao QA, Wang D, Wang J, Zhang J, Ren F, Chen R, Wang ZW;
PI Xue AJ, Yang Y, Wejhrman T, Goodrich R;
XX
DR WPI; 2001-476283/51.
DR P-PSDB; AAM78377.
XX
PT Nucleic acids encoding polypeptides with cytokine-like activities, useful
PT in diagnosis and gene therapy.
XX
PS Claim 1; Page 638-642; 6221pp; English.
XX
CC The invention relates to polynucleotides (AAK51456-AAK53435) and the
CC encoded polypeptides (AAM78323-AAM80302) that exhibit activity elating to
CC cytokine, cell proliferation or cell differentiation or which may induce
CC production of other cytokines in other cell populations. The
CC polynucleotides and polypeptides are useful in gene therapy, vaccines or
CC peptide therapy. The polypeptides have various cytokine-like activities,
CC e.g. stem cell growth factor activity, haematopoiesis regulating
CC activity, tissue growth factor activity, immunomodulatory activity and
CC activin/inhibin activity and may be useful in the diagnosis and/or
CC treatment of cancer, leukaemia, nervous system disorders, arthritis and
CC inflammation. Note: Records for SEQ ID NO 2110 (AAK52581), 2111
CC (AAK52582) and 3666 (AAM80020) are omitted as the relevant pages from the
CC sequence listing were missing at the time of publication
XX
SQ Sequence 5230 BP; 1285 A; 1255 C; 1330 G; 1360 T; 0 U; 0 Other;

Query Match 96.4%; Score 1460.2; DB 4; Length 5230;
Best Local Similarity 99.8%; Pred. No. 0;
Matches 1462; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

```
Qy         50 GCATGGGGCGCTTTGGGAACCAGGCCGATCACTTCTTGGGCTCTCTGGCATTTGCAAAGC 109
              ||  ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        206 GCCAGGGGCGCTTTGGGAACCAGGCCGATCACTTCTTGGGCTCTCTGGCATTTGCAAAGC 265

Qy        110 TGCTAAACCGTACCTTGGCTGTCCCTCCTTGGATTGAGTACCAGCATCACAAGCCTCCTT 169
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        266 TGCTAAACCGTACCTTGGCTGTCCCTCCTTGGATTGAGTACCAGCATCACAAGCCTCCTT 325

Qy        170 TCACCAACCTCCATGTGTCCTACCAGAAGTACTTCAAGCTGGAGCCCCTCCAGGCTTACC 229
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        326 TCACCAACCTCCATGTGTCCTACCAGAAGTACTTCAAGCTGGAGCCCCTCCAGGCTTACC 385

Qy        230 ATCGGGTCATCAGCTTGGAGGATTTCATGGAGAAGCTGGCACCCACCCACTGGCCCCCTG 289
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        386 ATCGGGTCATCAGCTTGGAGGATTTCATGGAGAAGCTGGCACCCACCCACTGGCCCCCTG 445

Qy        290 AGAAGCGGGTGGCATACTGCTTTGAGGTGGCAGCCCAGCGAAGCCCAGATAAGAAGACGT 349
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        446 AGAAGCGGGTGGCATACTGCTTTGAGGTGGCAGCCCAGCGAAGCCCAGATAAGAAGACGT 505

Qy        350 GCCCCATGAAGGAAGGAAACCCCTTTGGCCCATTCTGGGATCAGTTTCATGTGAGTTTCA 409
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        506 GCCCCATGAAGGAAGGAAACCCCTTTGGCCCATTCTGGGATCAGTTTCATGTGAGTTTCA 565

Qy        410 ACAAGTCGGAGCTTTTTACAGGCATTTCCTTCAGTGCTTCCTACAGAGAACAATGGAGCC 469
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        566 ACAAGTCGGAGCTTTTTACAGGCATTTCCTTCAGTGCTTCCTACAGAGAACAATGGAGCC 625

Qy        470 AGAGATTTTCTCCAAAGGAACATCCGGTGCTTGCCCTGCCAGGAGCCCCAGCCCAGTTCC 529
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        626 AGAGATTTTCTCCAAAGGAACATCCGGTGCTTGCCCTGCCAGGAGCCCCAGCCCAGTTCC 685

Qy        530 CCGTCCTAGAAGAACACAGGCCACTACAGAAGTACATGGTATGGTCAGACGAAATGGTGA 589
              ||||||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        686 CCGTCCTAGAGGAACACAGGCCACTACAGAAGTACATGGTATGGTCAGACGAAATGGTGA 745

Qy        590 AGACGGGAGAGGCCCAGATTCATGCCCACCTTGTCCGGCCCTATGTGGGCATTCATCTGC 649
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db        746 AGACGGGAGAGGCCCAGATTCATGCCCACCTTGTCCGGCCCTATGTGGGCATTCATCTGC 805

Qy        650 GCATTGGCTCTGACTGGAAGAACGCCTGTGCCATGCTGAAGGACGGGACTGCAGGCTCGC 709
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        806 GCATTGGCTCTGACTGGAAGAACGCCTGTGCCATGCTGAAGGACGGGACTGCAGGCTCGC 865

Qy        710 ACTTCATGGCCTCTCCGCAGTGTGTGGGCTACAGCCGCAGCACAGCGGCCCCCCTCACGA 769
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        866 ACTTCATGGCCTCTCCGCAGTGTGTGGGCTACAGCCGCAGCACAGCGGCCCCCCTCACGA 925

Qy        770 TGACTATGTGCCTGCCTGACCTGAAGGAGATCCAGAGGGCTGTGAAGCTCTGGGTGAGGT 829
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        926 TGACTATGTGCCTGCCTGACCTGAAGGAGATCCAGAGGGCTGTGAAGCTCTGGGTGAGGT 985

Qy        830 CGCTGGATGCCCAGTCGGTCTACGTTGCTACTGATTCCGAGAGTTATGTGCCTGAGCTCC 889
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        986 CGCTGGATGCCCAGTCGGTCTACGTTGCTACTGATTCCGAGAGTTATGTGCCTGAGCTCC 1045

Qy        890 AACAGCTCTTCAAAGGGAAGGTGAAGGTGGTGAGCCTGAAGCCTGAGGTGGCCCAGGTCG 949
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1046 AACAGCTCTTCAAAGGGAAGGTGAAGGTGGTGAGCCTGAAGCCTGAGGTGGCCCAGGTCG 1105

Qy        950 ACCTGTACATCCTCGGCCAAGCCGACCACTTTATTGGCAACTGTGTCTCCTCCTTCACTG 1009
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1106 ACCTGTACATCCTCGGCCAAGCCGACCACTTTATTGGCAACTGTGTCTCCTCCTTCACTG 1165

Qy       1010 CCTTTGTGAAGCGGGAGCGGGACCTCCAGGGGAGGCCGTCTTCTTTCTTCGGCATGGACA 1069
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1166 CCTTTGTGAAGCGGGAGCGGGACCTCCAGGGGAGGCCGTCTTCTTTCTTCGGCATGGACA 1225

Qy       1070 GGCCCCCTAAGCTGCGGGACGAGTTCTGATTCTGGCCGGAGCACCAGACCCTCTGATCCT 1129
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1226 GGCCCCCTAAGCTGCGGGACGAGTTCTGATTCTGGCCGGAGCACCAGACCCTCTGATCCT 1285

Qy       1130 GGAGGGACCAGAGTCTGAGCTGGTCCTTCCAGCCAGGCCTGGCAGCCAGAGGTGCTCCGG 1189
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1286 GGAGGGACCAGAGTCTGAGCTGGTCCTTCCAGCCAGGCCTGGCAGCCAGAGGTGCTCCGG 1345

Qy       1190 GATTGCAAACTCCTCTTCTCACCTGCCAAAGATGGAGAAGAGTGCCAGGGACCCCTCAAG 1249
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1346 GATTGCAAACTCCTCTTCTCACCTGCCAAAGATGGAGAAGAGTGCCAGGGACCCCTCAAG 1405

Qy       1250 GAGGGAGACGCTCCATATCCCAGGGCATAGGACTTGCAGGTTCCTAGGAGCAGGAGCATC 1309
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1406 GAGGGAGACGCTCCATATCCCAGGGCATAGGACTTGCAGGTTCCTAGGAGCAGGAGCATC 1465

Qy       1310 TCCCATCGCACGTGCTTTCTGCTCTTCTGGGAATTTCTCACACTGGCAAAGCAGTCCAGC 1369
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1466 TCCCATCGCACGTGCTTTCTGCTCTTCTGGGAATTTCTCACACTGGCAAAGCAGTCCAGC 1525

Qy       1370 CTCCGTCTTCTGGTCCACTCTGCTCTGAGCAGCCTGGGATGCTGAACTCTTCAGAGAGAT 1429
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1526 CTCCGTCTTCTGGTCCACTCTGCTCTGAGCAGCCTGGGATGCTGAACTCTTCAGAGAGAT 1585

Qy       1430 TTTTTTATAGAGAGATTTCTATAATTTTGATACAAGGTCATGACTATCCTAGAACTCTCT 1489
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1586 TTTTTTATAGAGAGATTTCTATAATTTTGATACAAGGTCATGACTATCCTAGAACTCTCT 1645

Qy       1490 GTGGTTTTTGAAAATCATTGAATTC 1514
              |||||||||||||||||||||||||
Db       1646 GTGGTTTTTGAAAATCATTGAATTC 1670


RESULT 6
AAK52494
ID   AAK52494 standard; cDNA; 4850 BP.
XX
AC   AAK52494;
XX
DT   06-NOV-2001  (first entry)
XX
DE   Human polynucleotide SEQ ID NO 2023.
XX
KW   Human; cytokine; cell proliferation; cell differentiation; gene therapy;
KW   vaccine; peptide therapy; stem cell growth factor; haematopoiesis;
KW   tissue growth factor; immunomodulatory; cancer; leukaemia;
KW   nervous system disorder; arthritis; inflammation; ss.
XX
OS   Homo sapiens.
XX
PN   WO200157190-A2.
XX
PD   09-AUG-2001.
XX
PF   05-FEB-2001; 2001WO-US004098.
XX
PR   03-FEB-2000; 2000US-00496914.
PR   27-APR-2000; 2000US-00560875.
PR   20-JUN-2000; 2000US-00598075.
PR   19-JUL-2000; 2000US-00620325.
PR   01-SEP-2000; 2000US-00654936.
PR   15-SEP-2000; 2000US-00663561.
PR   20-OCT-2000; 2000US-00693325.
PR   30-NOV-2000; 2000US-00728422.
XX
PA   (HYSE-) HYSEQ INC.
XX
PI   Tang YT,  Liu C,  Drmanac RT,  Asundi V,  Zhou P,  Xu C,  Cao Y;
PI   Ma Y,  Zhao QA,  Wang D,  Wang J,  Zhang J,  Ren F,  Chen R,  Wang ZW;
PI   Xue AJ,  Yang Y,  Wejhrman T,  Goodrich R;
XX
DR   WPI; 2001-476283/51.
DR   P-PSDB; AAM79361.
XX
PT   Nucleic acids encoding polypeptides with cytokine-like activities, useful
PT   in diagnosis and gene therapy.
XX
PS   Claim 1; Page 4424-4425; 6221pp; English.
XX
CC   The invention relates to polynucleotides (AAK51456-AAK53435) and the
CC   encoded polypeptides (AAM78323-AAM80302) that exhibit activity elating to
CC   cytokine, cell proliferation or cell differentiation or which may induce
CC   production of other cytokines in other cell populations. The
CC   polynucleotides and polypeptides are useful in gene therapy, vaccines or
CC   peptide therapy. The polypeptides have various cytokine-like activities,
CC   e.g. stem cell growth factor activity, haematopoiesis regulating
CC   activity, tissue growth factor activity, immunomodulatory activity and
CC   activin/inhibin activity and may be useful in the diagnosis and/or
CC   treatment of cancer, leukaemia, nervous system disorders, arthritis and
CC   inflammation. Note: Records for SEQ ID NO 2110 (AAK52581), 2111
CC   (AAK52582) and 3666 (AAM80020) are omitted as the relevant pages from the
CC   sequence listing were missing at the time of publication
XX
SQ   Sequence 4850 BP; 1142 A; 1218 C; 1244 G; 1246 T; 0 U; 0 Other;

  Query Match           96.1%;  Score 1455.4;  DB 4;  Length 4850;
  Best Local Similarity  99.6%;  Pred. No. 0;
  Matches 1459;  Conservative    0;  Mismatches    6;  Indels    0;  Gaps    0;

Qy         50 GCATGGGGCGCTTTGGGAACCAGGCCGATCACTTCTTGGGCTCTCTGGCATTTGCAAAGC 109
              ||  ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        206 GCCAGGGGCGCTTTGGGAACCAGGCCGATCACTTCTTGGGCTCTCTGGCATTTGCAAAGC 265

Qy        110 TGCTAAACCGTACCTTGGCTGTCCCTCCTTGGATTGAGTACCAGCATCACAAGCCTCCTT 169
              ||||  ||||||  ||||||||||||||  ||||||||||||||||||||||||||||||
Db        266 TGCTGAACCGTTCCTTGGCTGTCCCTTCTTGGATTGAGTACCAGCATCACAAGCCTCCTT 325

Qy        170 TCACCAACCTCCATGTGTCCTACCAGAAGTACTTCAAGCTGGAGCCCCTCCAGGCTTACC 229
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        326 TCACCAACCTCCATGTGTCCTACCAGAAGTACTTCAAGCTGGAGCCCCTCCAGGCTTACC 385

Qy        230 ATCGGGTCATCAGCTTGGAGGATTTCATGGAGAAGCTGGCACCCACCCACTGGCCCCCTG 289
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        386 ATCGGGTCATCAGCTTGGAGGATTTCATGGAGAAGCTGGCACCCACCCACTGGCCCCCTG 445

Qy        290 AGAAGCGGGTGGCATACTGCTTTGAGGTGGCAGCCCAGCGAAGCCCAGATAAGAAGACGT 349
```

```
Db        446 AGAAGCGGGTGGCATACTGCTTTGAGGTGGCAGCCCAGCGAAGCCCAGATAAGAAGACGT 505

Qy        350 GCCCCATGAAGGAAGGAAACCCCTTTGGCCCATTCTGGGATCAGTTTCATGTGAGTTTCA 409

Db        506 GCCCCATGAAGGAAGGAAACCCCTTTGGCCCATTCTGGGATCAGTTTCATGTGAGTTTCA 565

Qy        410 ACAAGTCGGAGCTTTTTACAGGCATTTCCTTCAGTGCTTCCTACAGAGAACAATGGAGCC 469

Db        566 ACAAGTCGGAGCTTTTTACAGGCATTTCCTTCAGTGCTTCCTACAGAGAACAATGGAGCC 625

Qy        470 AGAGATTTTCTCCAAAGGAACATCCGGTGCTTGCCCTGCCAGGAGCCCCAGCCCAGTTCC 529

Db        626 AGAGATTTTCTCCAAAGGAACATCCGGTGCTTGCCCTGCCAGGAGCCCCAGCCCAGTTCC 685

Qy        530 CCGTCCTAGAAGAACACAGGCCACTACAGAAGTACATGGTATGGTCAGACGAAATGGTGA 589

Db        686 CCGTCCTAGAGGAACACAGGCCACTACAGAAGTACATGGTATGGTCAGACGAAATGGTGA 745

Qy        590 AGACGGGAGAGGCCCAGATTCATGCCCACCTTGTCCGGCCCTATGTGGGCATTCATCTGC 649

Db        746 AGACGGGAGAGGCCCAGATTCATGCCCACCTTGTCCGGCCCTATGTGGGCATTCATCTGC 805

Qy        650 GCATTGGCTCTGACTGGAAGAACGCCTGTGCCATGCTGAAGGACGGGACTGCAGGCTCGC 709

Db        806 GCATTGGCTCTGACTGGAAGAACGCCTGTGCCATGCTGAAGGACGGGACTGCAGGCTCGC 865

Qy        710 ACTTCATGGCCTCTCCGCAGTGTGTGGGCTACAGCCGCAGCACAGCGGCCCCCCTCACGA 769

Db        866 ACTTCATGGCCTCTCCGCAGTGTGTGGGCTACAGCCGCAGCACAGCGGCCCCCCTCACGA 925

Qy        770 TGACTATGTGCCTGCCTGACCTGAAGGAGATCCAGAGGGCTGTGAAGCTCTGGGTGAGGT 829

Db        926 TGACTATGTGCCTGCCTGACCTGAAGGAGATCCAGAGGGCTGTGAAGCTCTGGGTGAGGT 985

Qy        830 CGCTGGATGCCCAGTCGGTCTACGTTGCTACTGATTCCGAGAGTTATGTGCCTGAGCTCC 889

Db        986 CGCTGGATGCCCAGTCGGTCTACGTTGCTACTGATTCCGAGAGTTATGTGCCTGAGCTCC 1045

Qy        890 AACAGCTCTTCAAAGGGAAGGTGAAGGTGGTGAGCCTGAAGCCTGAGGTGGCCCAGGTCG 949

Db       1046 AACAGCTCTTCAAAGGGAAGGTGAAGGTGGTGAGCCTGAAGCCTGAGGTGGCCCAGGTCG 1105

Qy        950 ACCTGTACATCCTCGGCCAAGCCGACCACTTTATTGGCAACTGTGTCTCCTCCTTCACTG 1009

Db       1106 ACCTGTACATCCTCGGCCAAGCCGACCACTTTATTGGCAACTGTGTCTCCTCCTTCACTG 1165

Qy       1010 CCTTTGTGAAGCGGGAGCGGGACCTCCAGGGGAGGCCGTCTTCTTTCTTCGGCATGGACA 1069

Db       1166 CCTTTGTGAAGCGGGAGCGGGACCTCCAGGGGAGGCCGTCTTCTTTCTTCGGCATGGACA 1225

Qy       1070 GGCCCCCTAAGCTGCGGGACGAGTTCTGATTCTGGCCGGAGCACCAGACCCTCTGATCCT 1129

Db       1226 GGCCCCCTAAGCTGCGGGACGAGTTCTGATTCTGGCCGGAGCACCAGACCCTCTGATCCT 1285

Qy       1130 GGAGGGACCAGAGTCTGAGCTGGTCCTTCCAGCCAGGCCTGGCAGCCAGAGGTGCTCCGG 1189

Db       1286 GGAGGGACCAGAGTCTGAGCTGGTCCTTCCAGCCAGGCCTGGCAGCCAGAGGTGCTCCGG 1345

Qy       1190 GATTGCAAACTCCTCTTCTCACCTGCCAAAGATGGAGAAGAGTGCCAGGGACCCCTCAAG 1249

Db       1346 GATTGCAAACTCCTCTTCTCACCTGCCAAAGATGGAGAAGAGTGCCAGGGACCCCTCAAG 1405

Qy       1250 GAGGGAGACGCTCCATATCCCAGGGCATAGGACTTGCAGGTTCCTAGGAGCAGGAGCATC 1309

Db       1406 GAGGGAGACGCTCCATATCCCAGGGCATAGGACTTGCAGGTTCCTAGGAGCAGGAGCATC 1465

Qy       1310 TCCCATCGCACGTGCTTTCTGCTCTTCTGGGAATTTCTCACACTGGCAAAGCAGTCCAGC 1369

Db       1466 TCCCATCGCACGTGCTTTCTGCTCTTCTGGGAATTTCTCACACTGGCAAAGCAGTCCAGC 1525

Qy       1370 CTCCGTCTTCTGGTCCACTCTGCTCTGAGCAGCCTGGGATGCTGAACTCTTCAGAGAGAT 1429

Db       1526 CTCCGTCTTCTGGTCCACTCTGCTCTGAGCAGCCTGGGATGCTGAACTCTTCAGAGAGAT 1585

Qy       1430 TTTTTTATAGAGAGATTTCTATAATTTTGATACAAGGTCATGACTATCCTAGAACTCTCT 1489

Db       1586 TTTTTTATAGAGAGATTTCTATAATTTTGATACAAGGTCATGACTATCCTAGAACTCTCT 1645

Qy       1490 GTGGTTTTTGAAAATCATTGAATTC 1514

Db       1646 GTGGTTTTTGAAAATCATTGAATTC 1670
```

RESULT 7
AAV65634
ID AAV65634 standard; cDNA; 5009 BP.
XX
AC AAV65634;
XX
DT 16-DEC-1998 (first entry)
XX
DE First EcoR1 nucleotide fragment of human KIAA0180.
XX
KW O-fucosyltransferase; epidermal growth factor; EGF; glycosylation;
KW O-fucose; inhibitor; sensory neuron; retinal neuron; human; heart; ss.
XX
OS Homo sapiens.
XX
PN WO9833924-A1.
XX
PD 06-AUG-1998.
XX
PF 17-DEC-1997; 97WO-US023401.
XX
PR 31-JAN-1997; 97US-00792498.
PR 26-NOV-1997; 97US-00978741.
XX
PA (GETH) GENENTECH INC.
XX
PI Wang Y, Spellman MW;
XX
DR WPI; 1998-437477/37.
DR P-PSDB; AAW80577.
XX
PT Human O-fucosyl:transferase able to glycosylate epidermal growth factor
PT domains - useful for diagnosis and treatment of diseases involving
PT overexpression of the enzyme.
XX
PS Example 1; Fig 11; 90pp; English.
XX
CC This represents a first EcoR1 nucleotide fragment of human KIAA0180. This
CC 5009 basepairs partial cDNA encodes for a protein of unknown function
CC from myeblast cel line KG-1. The invention provides a human heart O-
CC fucosyltransferase that can glycosylate an epidermal growth factor (EGF)
CC domain of a polypeptide with an activated O-fucose residue. Inhibitors of
CC O-fucosyltransferase, e.g. mutants with increased affinity for the EFG
CC domains, are used in diagnosis and treatment of conditions associated
CC with overexpression of O-fucosyltransferase, to promote survival of
CC sensory (retinal) neurons. Probes based on EGF domain polypeptide are
CC used to detect gene amplification and expression. The expression can also
CC be determined at the protein level using antibodies specific for O-
CC fucosyltransferase
XX
SQ Sequence 5009 BP; 1235 A; 1195 C; 1268 G; 1311 T; 0 U; 0 Other;

Query Match 95.6%; Score 1447.4; DB 2; Length 5009;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 1448; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```
Qy         66 GAACCAGGCCGATCACTTCTTGGGCTCTCTGGCATTTGCAAAGCTGCTAAACCGTACCTT 125

Db          1 GAACCAGGCCGATCACTTCTTGGGCTCTCTGGCATTTGCAAAGCTGCTAAACCGTACCTT 60

Qy        126 GGCTGTCCCTCCTTGGATTGAGTACCAGCATCACAAGCCTCCTTTCACCAACCTCCATGT 185
```

```
Db        61 GGCTGTCCCTCCTTGGATTGAGTACCAGCATCACAAGCCTCCTTTCACCAACCTCCATGT 120
Qy       186 GTCCTACCAGAAGTACTTCAAGCTGGAGCCCCTCCAGGCTTACCATCGGGTCATCAGCTT 245
Db       121 GTCCTACCAGAAGTACTTCAAGCTGGAGCCCCTCCAGGCTTACCATCGGGTCATCAGCTT 180
Qy       246 GGAGGATTTCATGGAGAAGCTGGCACCCACCCACTGGCCCCCTGAGAAGCGGGTGGCATA 305
Db       181 GGAGGATTTCATGGAGAAGCTGGCACCCACCCACTGGCCCCCTGAGAAGCGGGTGGCATA 240
Qy       306 CTGCTTTGAGGTGGCAGCCCAGCGAAGCCCAGATAAGAAGACGTGCCCCATGAAGGAAGG 365
Db       241 CTGCTTTGAGGTGGCAGCCCAGCGAAGCCCAGATAAGAAGACGTGCCCCATGAAGGAAGG 300
Qy       366 AAACCCCTTTGGCCCATTCTGGGATCAGTTTCATGTGAGTTTCAACAAGTCGGAGCTTTT 425
Db       301 AAACCCCTTTGGCCCATTCTGGGATCAGTTTCATGTGAGTTTCAACAAGTCGGAGCTTTT 360
Qy       426 TACAGGCATTTCCTTCAGTGCTTCCTACAGAGAACAATGGAGCCAGAGATTTTCTCCAAA 485
Db       361 TACAGGCATTTCCTTCAGTGCTTCCTACAGAGAACAATGGAGCCAGAGATTTTCTCCAAA 420
Qy       486 GGAACATCCGGTGCTTGCCCTGCCAGGAGCCCCAGCCCAGTTCCCCGTCCTAGAAGAACA 545
Db       421 GGAACATCCGGTGCTTGCCCTGCCAGGAGCCCCAGCCCAGTTCCCCGTCCTAGAGGAACA 480
Qy       546 CAGGCCACTACAGAAGTACATGGTATGGTCAGACGAAATGGTGAAGACGGGAGAGGCCCA 605
Db       481 CAGGCCACTACAGAAGTACATGGTATGGTCAGACGAAATGGTGAAGACGGGAGAGGCCCA 540
Qy       606 GATTCATGCCCACCTTGTCCGGCCCTATGTGGGCATTCATCTGCGCATTGGCTCTGACTG 665
Db       541 GATTCATGCCCACCTTGTCCGGCCCTATGTGGGCATTCATCTGCGCATTGGCTCTGACTG 600
Qy       666 GAAGAACGCCTGTGCCATGCTGAAGGACGGGACTGCAGGCTCGCACTTCATGGCCTCTCC 725
Db       601 GAAGAACGCCTGTGCCATGCTGAAGGACGGGACTGCAGGCTCGCACTTCATGGCCTCTCC 660
Qy       726 GCAGTGTGTGGGCTACAGCCGCAGCACAGCGGCCCCCCTCACGATGACTATGTGCCTGCC 785
Db       661 GCAGTGTGTGGGCTACAGCCGCAGCACAGCGGCCCCCCTCACGATGACTATGTGCCTGCC 720
Qy       786 TGACCTGAAGGAGATCCAGAGGGCTGTGAAGCTCTGGGTGAGGTCGCTGGATGCCCAGTC 845
Db       721 TGACCTGAAGGAGATCCAGAGGGCTGTGAAGCTCTGGGTGAGGTCGCTGGATGCCCAGTC 780
Qy       846 GGTCTACGTTGCTACTGATTCCGAGAGTTATGTGCCTGAGCTCCAACAGCTCTTCAAAGG 905
Db       781 GGTCTACGTTGCTACTGATTCCGAGAGTTATGTGCCTGAGCTCCAACAGCTCTTCAAAGG 840
Qy       906 GAAGGTGAAGGTGGTGAGCCTGAAGCCTGAGGTGGCCCAGGTCGACCTGTACATCCTCGG 965
Db       841 GAAGGTGAAGGTGGTGAGCCTGAAGCCTGAGGTGGCCCAGGTCGACCTGTACATCCTCGG 900
Qy       966 CCAAGCCGACCACTTTATTGGCAACTGTGTCTCCTCCTTCACTGCCTTTGTGAAGCGGGA 1025
Db       901 CCAAGCCGACCACTTTATTGGCAACTGTGTCTCCTCCTTCACTGCCTTTGTGAAGCGGGA 960
Qy      1026 GCGGGACCTCCAGGGGAGGCCGTCTTCTTTCTTCGGCATGGACAGGCCCCCTAAGCTGCG 1085
Db       961 GCGGGACCTCCAGGGGAGGCCGTCTTCTTTCTTCGGCATGGACAGGCCCCCTAAGCTGCG 1020
Qy      1086 GGACGAGTTCTGATTCTGGCCGGAGCACCAGACCCTCTGATCCTGGAGGGACCAGAGTCT 1145
Db       1021 GGACGAGTTCTGATTCTGGCCGGAGCACCAGACCCTCTGATCCTGGAGGGACCAGAGTCT 1080
Qy      1146 GAGCTGGTCCTTCCAGCCAGGCCTGGCAGCCAGAGGTGCTCCGGGATTGCAAACTCCTCT 1205
Db       1081 GAGCTGGTCCTTCCAGCCAGGCCTGGCAGCCAGAGGTGCTCCGGGATTGCAAACTCCTCT 1140
Qy      1206 TCTCACCTGCCAAAGATGGAGAAGAGTGCCAGGGACCCCTCAAGGAGGGAGACGCTCCAT 1265
Db       1141 TCTCACCTGCCAAAGATGGAGAAGAGTGCCAGGGACCCCTCAAGGAGGGAGACGCTCCAT 1200
Qy      1266 ATCCCAGGGCATAGGACTTGCAGGTTCCTAGGAGCAGGAGCATCTCCCATCGCACGTGCT 1325
Db       1201 ATCCCAGGGCATAGGACTTGCAGGTTCCTAGGAGCAGGAGCATCTCCCATCGCACGTGCT 1260
Qy      1326 TTCTGCTCTTCTGGGAATTTCTCACACTGGCAAAGCAGTCCAGCCTCCGTCTTCTGGTCC 1385
Db       1261 TTCTGCTCTTCTGGGAATTTCTCACACTGGCAAAGCAGTCCAGCCTCCGTCTTCTGGTCC 1320
Qy      1386 ACTCTGCTCTGAGCAGCCTGGGATGCTGAACTCTTCAGAGAGATTTTTTTATAGAGAGAT 1445
Db       1321 ACTCTGCTCTGAGCAGCCTGGGATGCTGAACTCTTCAGAGAGATTTTTTTATAGAGAGAT 1380
Qy      1446 TTCTATAATTTTGATACAAGGTCATGACTATCCTAGAACTCTCTGTGGTTTTTGAAAATC 1505
Db       1381 TTCTATAATTTTGATACAAGGTCATGACTATCCTAGAACTCTCTGTGGTTTTTGAAAATC 1440
Qy      1506 ATTGAATTC 1514
Db       1441 ATTGAATTC 1449
```

RESULT 8
AAD07629
ID AAD07629 standard; cDNA; 915 BP.
XX
AC AAD07629;
XX
DT 10-AUG-2001 (first entry)
XX
DE Human secreted protein-encoding gene 18 cDNA clone HMWEY26, SEQ ID NO:69.
XX
KW Human; secreted protein; proliferative disorder; cancer; tumour;
KW foetal abnormality; developmental abnormality; haematopoietic disorder;
KW immune system disorder; AIDS; autoimmune disease; rheumatoid arthritis;
KW inflammation; allergy; neurological disorder; Alzheimer's disease;
KW Parkinson's disease; cognitive disorder; schizophrenia; asthma;
KW skin disorder; psoriasis; sepsis; diabetes; atherosclerosis;
KW cardiovascular disorder; angiogenic disorder; kidney disorder;
KW gastrointestinal disorder; pregnancy-related disorder; gene therapy;
KW endocrine disorder; infection; wound healing; vulnerary; cell culture;
KW chemotaxis; food additive; binding partner identification; chromosome 3;
KW ss.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT CDS 1. .498
FT /*tag= a
FT /product= "Human secreted protein precursor"
FT /transl_except= (pos:4. .6, aa:Xaa)
FT /transl_except= (pos:10. .12, aa:Xaa)
FT /transl_except= (pos:13. .15, aa:Xaa)
FT /transl_except= (pos:442. .444, aa:Xaa)
FT /note= "Xaa equals any of the naturally occurring L-amino
FT acids; CDS does not include start codon"
FT /partial
FT sig_peptide 1. .3
FT /*tag= b
FT mat_peptide 4. .495
FT /*tag= c
FT /product= "Mature human secreted protein"
XX
PN WO200132676-A1.
XX
PD 10-MAY-2001.
XX
PF 25-OCT-2000; 2000WO-US029365.
XX
PR 29-OCT-1999; 99US-0162237P.
PR 21-JUL-2000; 2000US-0219666P.
XX
PA (HUMA-) HUMAN GENOME SCI INC.

```
XX
PI   Ruben SM,  Komatsoulis GA,  Shi Y,  Olsen HS,  Soppet DR;
XX
DR   WPI; 2001-328773/34.
DR   P-PSDB; AAE03110.
XX
PT   Nucleic acids encoding 25 human secreted polypeptides, useful for
PT   preventing, diagnosing and/or treating e.g. Gaucher's disease,
PT   Alzheimer's disease, Scimitar syndrome, Creutzfeldt-Jacob disease,
PT   diabetes mellitus and multiple sclerosis.
XX
PS   Claim 1; Page 445; 546pp; English.
XX
CC   AAD07571-AAD07645 represent cDNAs corresponding to 25 human secreted
CC   protein genes, and AAE03052-AAE03126 represent the proteins they encode.
CC   AAE03127-AAE03150 represent human secreted protein fragments. The genes
CC   and their corresponding secreted proteins are useful for preventing,
CC   treating or ameliorating medical conditions, e.g., by protein or gene
CC   therapy. Pathological conditions can be diagnosed by determining the
CC   amount of the new protein in a sample or by determining the presence of
CC   mutations in the new genes. Specific uses are described for each of the
CC   25 genes, based on the tissues in which they are most highly expressed,
CC   and include developing products for the diagnosis or treatment of
CC   proliferative disorders, cancer, tumours, foetal and developmental
CC   abnormalities, haematopoietic disorders, diseases of the immune system,
CC   AIDS, autoimmune diseases (e.g., rheumatoid arthritis), inflammation,
CC   allergies, neurological disorders (e.g., Alzheimer's disease, Parkinson's
CC   disease), cognitive disorders, schizophrenia, asthma, skin disorders
CC   (e.g., psoriasis), sepsis, diabetes, atherosclerosis, cardiovascular
CC   disorders, angiogenic disorders, kidney disorders, gastrointestinal
CC   disorders, pregnancy-related disorders, endocrine disorders, and
CC   infections. The proteins can also be used to aid wound healing and
CC   epithelial cell proliferation, to prevent skin aging due to sunburn, to
CC   maintain organs before transplantation, for supporting cell culture of
CC   primary tissues, to regenerate tissues, to identify their cognate ligands
CC   or binding partners, and in chemotaxis, and can be used as a food
CC   additive or preservative to modify storage properties. Antibodies
CC   specific for a protein of the invention can be used in alleviating
CC   symptoms associated with the disorders mentioned above, and in diagnostic
CC   immunoassays e.g., radioimmunoassay or enzyme linked immunosorbent assay
CC   (ELISA). The present sequence represents a human secreted protein-
CC   encoding cDNA of the invention
XX
SQ   Sequence 915 BP; 209 A; 253 C; 247 G; 202 T; 0 U; 4 Other;

  Query Match             56.5%;  Score 854.8;  DB 4;  Length 915;
  Best Local Similarity   99.0%;  Pred. No. 1.4e-238;
  Matches  859;  Conservative    1;  Mismatches    8;  Indels    0;  Gaps    0;

Qy        616 CACCTTGTCCGGCCCTATGTGGGCATTCATCTGCGCATTGGCTCTGACTGGAAGAACGCC 675
              ||| ||| |||  ||||  ||||||||||||||||||||||||||||||||||||||||
Db         16 CACTTTGCCCGAACCTACCCGGGCATTCATCTGCGCATTGGCTCTGACTGGAAGAACGCC 75

Qy        676 TGTGCCATGCTGAAGGACGGGACTGCAGGCTCGCACTTCATGGCCTCTCCGCAGTGTGTG 735
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         76 TGTGCCATGCTGAAGGACGGGACTGCAGGCTCGCACTTCATGGCCTCTCCGCAGTGTGTG 135

Qy        736 GGCTACAGCCGCAGCACAGCGGCCCCCCTCACGATGACTATGTGCCTGCCTGACCTGAAG 795
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        136 GGCTACAGCCGCAGCACAGCGGCCCCCCTCACGATGACTATGTGCCTGCCTGACCTGAAG 195

Qy        796 GAGATCCAGAGGGCTGTGAAGCTCTGGGTGAGGTCGCTGGATGCCCAGTCGGTCTACGTT 855
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        196 GAGATCCAGAGGGCTGTGAAGCTCTGGGTGAGGTCGCTGGATGCCCAGTCGGTCTACGTT 255

Qy        856 GCTACTGATTCCGAGAGTTATGTGCCTGAGCTCCAACAGCTCTTCAAAGGGAAGGTGAAG 915
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        256 GCTACTGATTCCGAGAGTTATGTGCCTGAGCTCCAACAGCTCTTCAAAGGGAAGGTGAAG 315

Qy        916 GTGGTGAGCCTGAAGCCTGAGGTGGCCCAGGTCGACCTGTACATCCTCGGCCAAGCCGAC 975
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        316 GTGGTGAGCCTGAAGCCTGAGGTGGCCCAGGTCGACCTGTACATCCTCGGCCAAGCCGAC 375

Qy        976 CACTTTATTGGCAACTGTGTCTCCTCCTTCACTGCCTTTGTGAAGCGGGAGCGGGACCTC 1035
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        376 CACTTTATTGGCAACTGTGTCTCCTCCTTCACTGCCTTTGTGAAGCGGGAGCGGGACCTC 435

Qy       1036 CAGGGGAGGCCGTCTTCTTTCTTCGGCATGGACAGGCCCCCTAAGCTGCGGGACGAGTTC 1095
              |||||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        436 CAGGGGAGSCCGTCTTCTTTCTTCGGCATGGACAGGCCCCCTAAGCTGCGGGACGAGTTC 495

Qy       1096 TGATTCTGGCCGGAGCACCAGACCCTCTGATCCTGGAGGGACCAGAGTCTGAGCTGGTCC 1155
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        496 TGATTCTGGCCGGAGCACCAGACCCTCTGATCCTGGAGGGACCAGAGTCTGAGCTGGTCC 555

Qy       1156 TTCCAGCCAGGCCTGGCAGCCAGAGGTGCTCCGGGATTGCAAACTCCTCTTCTCACCTGC 1215
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        556 TTCCAGCCAGGCCTGGCAGCCAGAGGTGCTCCGGGATTGCAAACTCCTCTTCTCACCTGC 615

Qy       1216 CAAAGATGGAGAAGAGTGCCAGGGACCCCTCAAGGAGGGAGACGCTCCATATCCCAGGGC 1275
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        616 CAAAGATGGAGAAGAGTGCCAGGGACCCCTCAAGGAGGGAGACGCTCCATATCCCAGGGC 675

Qy       1276 ATAGGACTTGCAGGTTCCTAGGAGCAGGAGCATCTCCCATCGCACGTGCTTTCTGCTCTT 1335
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        676 ATAGGACTTGCAGGTTCCTAGGAGCAGGAGCATCTCCCATCGCACGTGCTTTCTGCTCTT 735

Qy       1336 CTGGGAATTTCTCACACTGGCAAAGCAGTCCAGCCTCCGTCTTCTGGTCCACTCTGCTCT 1395
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        736 CTGGGAATTTCTCACACTGGCAAAGCAGTCCAGCCTCCGTCTTCTGGTCCACTCTGCTCT 795

Qy       1396 GAGCAGCCTGGGATGCTGAACTCTTCAGAGAGATTTTTTTATAGAGAGATTTCTATAATT 1455
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        796 GAGCAGCCTGGGATGCTGAACTCTTCAGAGAGATTTTTTTATAGAGAGATTTCTATAATT 855

Qy       1456 TTGATACAAGGTCATGACTATCCTAGAA 1483
              |||||||||||||||||||||||| ||
Db        856 TTGATACAAGGTCATGACTATCCTAAAA 883


RESULT 9
AAK79915
ID   AAK79915 standard; DNA; 10331 BP.
XX
AC   AAK79915;
XX
DT   07-NOV-2001  (first entry)
XX
DE   Human immune/haematopoietic antigen genomic sequence SEQ ID NO:34727.
XX
KW   Human; immune; haematopoietic; immune/haematopoietic antigen; cancer;
KW   cytostatic; gene therapy; vaccine; metastasis; ds.
XX
OS   Homo sapiens.
XX
PN   WO200157182-A2.
XX
PD   09-AUG-2001.
XX
PF   17-JAN-2001; 2001WO-US001354.
XX
PR   31-JAN-2000; 2000US-0179065P.
PR   04-FEB-2000; 2000US-0180628P.
PR   24-FEB-2000; 2000US-0184664P.
PR   02-MAR-2000; 2000US-0186350P.
PR   16-MAR-2000; 2000US-0189874P.
PR   17-MAR-2000; 2000US-0190076P.
PR   18-APR-2000; 2000US-0198123P.
PR   19-MAY-2000; 2000US-0205515P.
PR   07-JUN-2000; 2000US-0209467P.
PR   28-JUN-2000; 2000US-0214886P.
PR   30-JUN-2000; 2000US-0215135P.
PR   07-JUL-2000; 2000US-0216647P.
PR   07-JUL-2000; 2000US-0216880P.
PR   11-JUL-2000; 2000US-0217487P.
PR   11-JUL-2000; 2000US-0217496P.
```

PR 14-JUL-2000; 2000US-0218290P.
PR 26-JUL-2000; 2000US-0220963P.
PR 26-JUL-2000; 2000US-0220964P.
PR 14-AUG-2000; 2000US-0224518P.
PR 14-AUG-2000; 2000US-0224519P.
PR 14-AUG-2000; 2000US-0225213P.
PR 14-AUG-2000; 2000US-0225214P.
PR 14-AUG-2000; 2000US-0225266P.
PR 14-AUG-2000; 2000US-0225267P.
PR 14-AUG-2000; 2000US-0225268P.
PR 14-AUG-2000; 2000US-0225270P.
PR 14-AUG-2000; 2000US-0225447P.
PR 14-AUG-2000; 2000US-0225757P.
PR 14-AUG-2000; 2000US-0225758P.
PR 14-AUG-2000; 2000US-0225759P.
PR 18-AUG-2000; 2000US-0226279P.
PR 22-AUG-2000; 2000US-0226681P.
PR 22-AUG-2000; 2000US-0226868P.
PR 22-AUG-2000; 2000US-0227182P.
PR 23-AUG-2000; 2000US-0227009P.
PR 30-AUG-2000; 2000US-0228924P.
PR 01-SEP-2000; 2000US-0229287P.
PR 01-SEP-2000; 2000US-0229343P.
PR 01-SEP-2000; 2000US-0229344P.
PR 01-SEP-2000; 2000US-0229345P.
PR 05-SEP-2000; 2000US-0229509P.
PR 05-SEP-2000; 2000US-0229513P.
PR 06-SEP-2000; 2000US-0230437P.
PR 06-SEP-2000; 2000US-0230438P.
PR 08-SEP-2000; 2000US-0231242P.
PR 08-SEP-2000; 2000US-0231243P.
PR 08-SEP-2000; 2000US-0231244P.
PR 08-SEP-2000; 2000US-0231413P.
PR 08-SEP-2000; 2000US-0231414P.
PR 08-SEP-2000; 2000US-0232080P.
PR 08-SEP-2000; 2000US-0232081P.
PR 12-SEP-2000; 2000US-0231968P.
PR 14-SEP-2000; 2000US-0232397P.
PR 14-SEP-2000; 2000US-0232398P.
PR 14-SEP-2000; 2000US-0232399P.
PR 14-SEP-2000; 2000US-0232400P.
PR 14-SEP-2000; 2000US-0232401P.
PR 14-SEP-2000; 2000US-0233063P.
PR 14-SEP-2000; 2000US-0233064P.
PR 14-SEP-2000; 2000US-0233065P.
PR 21-SEP-2000; 2000US-0234223P.
PR 21-SEP-2000; 2000US-0234274P.
PR 25-SEP-2000; 2000US-0234997P.
PR 25-SEP-2000; 2000US-0234998P.
PR 26-SEP-2000; 2000US-0235484P.
PR 27-SEP-2000; 2000US-0235834P.
PR 27-SEP-2000; 2000US-0235836P.
PR 29-SEP-2000; 2000US-0236327P.
PR 29-SEP-2000; 2000US-0236367P.
PR 29-SEP-2000; 2000US-0236368P.
PR 29-SEP-2000; 2000US-0236369P.
PR 29-SEP-2000; 2000US-0236370P.
PR 02-OCT-2000; 2000US-0236802P.
PR 02-OCT-2000; 2000US-0237037P.
PR 02-OCT-2000; 2000US-0237038P.
PR 02-OCT-2000; 2000US-0237039P.
PR 02-OCT-2000; 2000US-0237040P.
PR 13-OCT-2000; 2000US-0239935P.
PR 13-OCT-2000; 2000US-0239937P.
PR 20-OCT-2000; 2000US-0240960P.
PR 20-OCT-2000; 2000US-0241221P.
PR 20-OCT-2000; 2000US-0241785P.
PR 20-OCT-2000; 2000US-0241786P.
PR 20-OCT-2000; 2000US-0241787P.
PR 20-OCT-2000; 2000US-0241808P.
PR 20-OCT-2000; 2000US-0241809P.
PR 20-OCT-2000; 2000US-0241826P.
PR 01-NOV-2000; 2000US-0244617P.
PR 08-NOV-2000; 2000US-0246474P.
PR 08-NOV-2000; 2000US-0246475P.
PR 08-NOV-2000; 2000US-0246476P.
PR 08-NOV-2000; 2000US-0246477P.
PR 08-NOV-2000; 2000US-0246478P.
PR 08-NOV-2000; 2000US-0246523P.
PR 08-NOV-2000; 2000US-0246524P.
PR 08-NOV-2000; 2000US-0246525P.
PR 08-NOV-2000; 2000US-0246526P.
PR 08-NOV-2000; 2000US-0246527P.
PR 08-NOV-2000; 2000US-0246528P.
PR 08-NOV-2000; 2000US-0246532P.
PR 08-NOV-2000; 2000US-0246609P.
PR 08-NOV-2000; 2000US-0246610P.
PR 08-NOV-2000; 2000US-0246611P.
PR 08-NOV-2000; 2000US-0246613P.
PR 17-NOV-2000; 2000US-0249207P.
PR 17-NOV-2000; 2000US-0249208P.
PR 17-NOV-2000; 2000US-0249209P.
PR 17-NOV-2000; 2000US-0249210P.
PR 17-NOV-2000; 2000US-0249211P.
PR 17-NOV-2000; 2000US-0249212P.
PR 17-NOV-2000; 2000US-0249213P.
PR 17-NOV-2000; 2000US-0249214P.
PR 17-NOV-2000; 2000US-0249215P.
PR 17-NOV-2000; 2000US-0249216P.
PR 17-NOV-2000; 2000US-0249217P.
PR 17-NOV-2000; 2000US-0249218P.
PR 17-NOV-2000; 2000US-0249244P.
PR 17-NOV-2000; 2000US-0249245P.
PR 17-NOV-2000; 2000US-0249264P.
PR 17-NOV-2000; 2000US-0249265P.
PR 17-NOV-2000; 2000US-0249297P.
PR 17-NOV-2000; 2000US-0249299P.
PR 17-NOV-2000; 2000US-0249300P.
PR 01-DEC-2000; 2000US-0250160P.
PR 01-DEC-2000; 2000US-0250391P.
PR 05-DEC-2000; 2000US-0251030P.
PR 05-DEC-2000; 2000US-0251988P.
PR 05-DEC-2000; 2000US-0256719P.
PR 06-DEC-2000; 2000US-0251479P.
PR 08-DEC-2000; 2000US-0251856P.
PR 08-DEC-2000; 2000US-0251868P.
PR 08-DEC-2000; 2000US-0251869P.
PR 08-DEC-2000; 2000US-0251989P.
PR 08-DEC-2000; 2000US-0251990P.
PR 11-DEC-2000; 2000US-0254097P.
PR 05-JAN-2001; 2001US-0259678P.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Rosen CA, Barash SC, Ruben SM;
XX
DR WPI; 2001-483426/52.
XX
PT Nucleic acids encoding human immune/hematopoietic antigen polypeptides,
PT useful for preventing, diagnosing and/or treating cancers and metastasis.
XX
PS Disclosure; SEQ ID NO 34727; 3071pp + Sequence Listing; English.
XX
CC AAK54951 to AAK64702 encode the human immune/haematopoietic antigen (I)
CC amino acid sequences given in AAM82170 to AAM91921. (I) have cytostatic
CC activity, and can be used in gene therapy and vaccine production. (I)
CC proteins and polynucleotides may be used in the prevention, diagnosis and
CC treatment of diseases associated with inappropriate (I) expression. For
CC example, they may be used to treat disorders associated with decreased
CC expression by rectifying mutations or deletions in a patient's genome
CC that affect the activity of (I) by expressing inactive proteins or to
CC supplement the patients own production of (I). Additionally, (I)
CC polynucleotides may be used to produce the secreted (I), by inserting the
CC nucleic acids into a host cell and culturing the cell to express the
CC protein. (I) proteins and polynucleotides may be used to prevent,
CC diagnose and treat immune/haematopoietic-related diseases, especially

```
CC   cancers and cancer metastases of haematopoietic-derived cells. AAK64703
CC   to AAK87694 represent human immune/haematopoietic antigen genomic
CC   sequences from the present invention. AAK54942 to AAK54950 and AAM82169
CC   represent sequences used in the exemplification of the present invention
XX
SQ   Sequence 10331 BP; 2718 A; 2405 C; 2599 G; 2609 T; 0 U; 0 Other;

  Query Match             40.2%;  Score 608.6;  DB 4;  Length 10331;
  Best Local Similarity   98.6%;  Pred. No. 4e-166;
  Matches  614;  Conservative    0;  Mismatches    9;  Indels    0;  Gaps    0;

Qy        892 CAGCTCTTCAAAGGGAAGGTGAAGGTGGTGAGCCTGAAGCCTGAGGTGGCCCAGGTCGAC 951
              | |  ||||     | ||||||||||||||||||||||||||||||||||||||||||||
Db       6127 CTGTGCTTCTTCCTGCAGGTGAAGGTGGTGAGCCTGAAGCCTGAGGTGGCCCAGGTCGAC 6186

Qy        952 CTGTACATCCTCGGCCAAGCCGACCACTTTATTGGCAACTGTGTCTCCTCCTTCACTGCC 1011
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       6187 CTGTACATCCTCGGCCAAGCCGACCACTTTATTGGCAACTGTGTCTCCTCCTTCACTGCC 6246

Qy       1012 TTTGTGAAGCGGGAGCGGGACCTCCAGGGGAGGCCGTCTTCTTTCTTCGGCATGGACAGG 1071
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       6247 TTTGTGAAGCGGGAGCGGGACCTCCAGGGGAGGCCGTCTTCTTTCTTCGGCATGGACAGG 6306

Qy       1072 CCCCCTAAGCTGCGGGACGAGTTCTGATTCTGGCCGGAGCACCAGACCCTCTGATCCTGG 1131
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       6307 CCCCCTAAGCTGCGGGACGAGTTCTGATTCTGGCCGGAGCACCAGACCCTCTGATCCTGG 6366

Qy       1132 AGGGACCAGAGTCTGAGCTGGTCCTTCCAGCCAGGCCTGGCAGCCAGAGGTGCTCCGGGA 1191
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       6367 AGGGACCAGAGTCTGAGCTGGTCCTTCCAGCCAGGCCTGGCAGCCAGAGGTGCTCCGGGA 6426

Qy       1192 TTGCAAACTCCTCTTCTCACCTGCCAAAGATGGAGAAGAGTGCCAGGGACCCCTCAAGGA 1251
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       6427 TTGCAAACTCCTCTTCTCACCTGCCAAAGATGGAGAAGAGTGCCAGGGACCCCTCAAGGA 6486

Qy       1252 GGGAGACGCTCCATATCCCAGGGCATAGGACTTGCAGGTTCCTAGGAGCAGGAGCATCTC 1311
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       6487 GGGAGACGCTCCATATCCCAGGGCATAGGACTTGCAGGTTCCTAGGAGCAGGAGCATCTC 6546

Qy       1312 CCATCGCACGTGCTTTCTGCTCTTCTGGGAATTTCTCACACTGGCAAAGCAGTCCAGCCT 1371
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       6547 CCATCGCACGTGCTTTCTGCTCTTCTGGGAATTTCTCACACTGGCAAAGCAGTCCAGCCT 6606

Qy       1372 CCGTCTTCTGGTCCACTCTGCTCTGAGCAGCCTGGGATGCTGAACTCTTCAGAGAGATTT 1431
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       6607 CCGTCTTCTGGTCCACTCTGCTCTGAGCAGCCTGGGATGCTGAACTCTTCAGAGAGATTT 6666

Qy       1432 TTTTATAGAGAGATTTCTATAATTTTGATACAAGGTCATGACTATCCTAGAACTCTCTGT 1491
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       6667 TTTTATAGAGAGATTTCTATAATTTTGATACAAGGTCATGACTATCCTAGAACTCTCTGT 6726

Qy       1492 GGTTTTTGAAAATCATTGAATTC 1514
              |||||||||||||||||||||||
Db       6727 GGTTTTTGAAAATCATTGAATTC 6749


RESULT 10
AAL03344
ID   AAL03344 standard; DNA; 10331 BP.
XX
AC   AAL03344;
XX
DT   21-NOV-2001  (first entry)
XX
DE   Human reproductive system related antigen DNA SEQ ID NO: 6032.
XX
KW   Human; reproductive system related antigen; reproductive system disorder;
KW   cancer; gene therapy; ds.
XX
OS   Homo sapiens.
XX
PN   WO200155320-A2.
XX
PD   02-AUG-2001.
XX
PF   17-JAN-2001; 2001WO-US001339.
XX
PR   31-JAN-2000; 2000US-0179065P.
PR   04-FEB-2000; 2000US-0180628P.
PR   24-FEB-2000; 2000US-0184664P.
PR   02-MAR-2000; 2000US-0186350P.
PR   16-MAR-2000; 2000US-0189874P.
PR   17-MAR-2000; 2000US-0190076P.
PR   18-APR-2000; 2000US-0198123P.
PR   19-MAY-2000; 2000US-0205515P.
PR   07-JUN-2000; 2000US-0209467P.
PR   28-JUN-2000; 2000US-0214886P.
PR   30-JUN-2000; 2000US-0215135P.
PR   07-JUL-2000; 2000US-0216647P.
PR   07-JUL-2000; 2000US-0216880P.
PR   11-JUL-2000; 2000US-0217487P.
PR   11-JUL-2000; 2000US-0217496P.
PR   14-JUL-2000; 2000US-0218290P.
PR   26-JUL-2000; 2000US-0220963P.
PR   26-JUL-2000; 2000US-0220964P.
PR   14-AUG-2000; 2000US-0224518P.
PR   14-AUG-2000; 2000US-0224519P.
PR   14-AUG-2000; 2000US-0225213P.
PR   14-AUG-2000; 2000US-0225214P.
PR   14-AUG-2000; 2000US-0225266P.
PR   14-AUG-2000; 2000US-0225267P.
PR   14-AUG-2000; 2000US-0225268P.
PR   14-AUG-2000; 2000US-0225270P.
PR   14-AUG-2000; 2000US-0225447P.
PR   14-AUG-2000; 2000US-0225757P.
PR   14-AUG-2000; 2000US-0225758P.
PR   14-AUG-2000; 2000US-0225759P.
PR   18-AUG-2000; 2000US-0226279P.
PR   22-AUG-2000; 2000US-0226681P.
PR   22-AUG-2000; 2000US-0226868P.
PR   22-AUG-2000; 2000US-0227182P.
PR   23-AUG-2000; 2000US-0227009P.
PR   30-AUG-2000; 2000US-0228924P.
PR   01-SEP-2000; 2000US-0229287P.
PR   01-SEP-2000; 2000US-0229343P.
PR   01-SEP-2000; 2000US-0229344P.
PR   01-SEP-2000; 2000US-0229345P.
PR   05-SEP-2000; 2000US-0229509P.
PR   05-SEP-2000; 2000US-0229513P.
PR   06-SEP-2000; 2000US-0230437P.
PR   06-SEP-2000; 2000US-0230438P.
PR   08-SEP-2000; 2000US-0231242P.
PR   08-SEP-2000; 2000US-0231243P.
PR   08-SEP-2000; 2000US-0231244P.
PR   08-SEP-2000; 2000US-0231413P.
PR   08-SEP-2000; 2000US-0231414P.
PR   08-SEP-2000; 2000US-0232080P.
PR   08-SEP-2000; 2000US-0232081P.
PR   12-SEP-2000; 2000US-0231968P.
PR   14-SEP-2000; 2000US-0232397P.
PR   14-SEP-2000; 2000US-0232398P.
PR   14-SEP-2000; 2000US-0232399P.
PR   14-SEP-2000; 2000US-0232400P.
PR   14-SEP-2000; 2000US-0232401P.
PR   14-SEP-2000; 2000US-0233063P.
PR   14-SEP-2000; 2000US-0233064P.
PR   14-SEP-2000; 2000US-0233065P.
PR   21-SEP-2000; 2000US-0234223P.
PR   21-SEP-2000; 2000US-0234274P.
PR   25-SEP-2000; 2000US-0234997P.
PR   25-SEP-2000; 2000US-0234998P.
PR   26-SEP-2000; 2000US-0235484P.
PR   27-SEP-2000; 2000US-0235834P.
PR   27-SEP-2000; 2000US-0235836P.
PR   29-SEP-2000; 2000US-0236327P.
PR   29-SEP-2000; 2000US-0236367P.
```

```
PR   29-SEP-2000; 2000US-0236368P.
PR   29-SEP-2000; 2000US-0236369P.
PR   29-SEP-2000; 2000US-0236370P.
PR   02-OCT-2000; 2000US-0236802P.
PR   02-OCT-2000; 2000US-0237037P.
PR   02-OCT-2000; 2000US-0237038P.
PR   02-OCT-2000; 2000US-0237039P.
PR   02-OCT-2000; 2000US-0237040P.
PR   13-OCT-2000; 2000US-0239935P.
PR   13-OCT-2000; 2000US-0239937P.
PR   20-OCT-2000; 2000US-0240960P.
PR   20-OCT-2000; 2000US-0241221P.
PR   20-OCT-2000; 2000US-0241785P.
PR   20-OCT-2000; 2000US-0241786P.
PR   20-OCT-2000; 2000US-0241787P.
PR   20-OCT-2000; 2000US-0241808P.
PR   20-OCT-2000; 2000US-0241809P.
PR   20-OCT-2000; 2000US-0241826P.
PR   01-NOV-2000; 2000US-0244617P.
PR   08-NOV-2000; 2000US-0246474P.
PR   08-NOV-2000; 2000US-0246475P.
PR   08-NOV-2000; 2000US-0246476P.
PR   08-NOV-2000; 2000US-0246477P.
PR   08-NOV-2000; 2000US-0246478P.
PR   08-NOV-2000; 2000US-0246523P.
PR   08-NOV-2000; 2000US-0246524P.
PR   08-NOV-2000; 2000US-0246525P.
PR   08-NOV-2000; 2000US-0246526P.
PR   08-NOV-2000; 2000US-0246527P.
PR   08-NOV-2000; 2000US-0246528P.
PR   08-NOV-2000; 2000US-0246532P.
PR   08-NOV-2000; 2000US-0246609P.
PR   08-NOV-2000; 2000US-0246610P.
PR   08-NOV-2000; 2000US-0246611P.
PR   08-NOV-2000; 2000US-0246613P.
PR   17-NOV-2000; 2000US-0249207P.
PR   17-NOV-2000; 2000US-0249208P.
PR   17-NOV-2000; 2000US-0249209P.
PR   17-NOV-2000; 2000US-0249210P.
PR   17-NOV-2000; 2000US-0249211P.
PR   17-NOV-2000; 2000US-0249212P.
PR   17-NOV-2000; 2000US-0249213P.
PR   17-NOV-2000; 2000US-0249214P.
PR   17-NOV-2000; 2000US-0249215P.
PR   17-NOV-2000; 2000US-0249216P.
PR   17-NOV-2000; 2000US-0249217P.
PR   17-NOV-2000; 2000US-0249218P.
PR   17-NOV-2000; 2000US-0249244P.
PR   17-NOV-2000; 2000US-0249245P.
PR   17-NOV-2000; 2000US-0249264P.
PR   17-NOV-2000; 2000US-0249265P.
PR   17-NOV-2000; 2000US-0249297P.
PR   17-NOV-2000; 2000US-0249299P.
PR   17-NOV-2000; 2000US-0249300P.
PR   01-DEC-2000; 2000US-0250160P.
PR   01-DEC-2000; 2000US-0250391P.
PR   05-DEC-2000; 2000US-0251030P.
PR   05-DEC-2000; 2000US-0251988P.
PR   05-DEC-2000; 2000US-0256719P.
PR   06-DEC-2000; 2000US-0251479P.
PR   08-DEC-2000; 2000US-0251856P.
PR   08-DEC-2000; 2000US-0251868P.
PR   08-DEC-2000; 2000US-0251869P.
PR   08-DEC-2000; 2000US-0251989P.
PR   08-DEC-2000; 2000US-0251990P.
PR   11-DEC-2000; 2000US-0254097P.
PR   05-JAN-2001; 2001US-0259678P.
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
XX
PI   Rosen CA,  Barash SC,  Ruben SM;
XX
DR   WPI; 2001-465570/50.
XX
PT   Isolated nucleic acid molecule encoding a reproductive system antigen is
PT   used in preventing, treating or ameliorating a medical condition.
XX
PS   Disclosure; SEQ ID NO 6032; 1297pp + Sequence Listing; English.
XX
CC   The present invention provides the protein and coding sequences of a
CC   number of human reproductive system related antigens. These can be used
CC   in the prevention and treatment of reproductive system disorders,
CC   including cancer. The present sequence is a genomic sequence encoding a
CC   protein of the invention
XX
SQ   Sequence 10331 BP; 2718 A; 2405 C; 2599 G; 2609 T; 0 U; 0 Other;

  Query Match             40.2%;  Score 608.6;  DB 4;  Length 10331;
  Best Local Similarity   98.6%;  Pred. No. 4e-166;
  Matches  614;  Conservative    0;  Mismatches    9;  Indels    0;  Gaps    0;

Qy        892 CAGCTCTTCAAAGGGAAGGTGAAGGTGGTGAGCCTGAAGCCTGAGGTGGCCCAGGTCGAC 951
              | |  ||||     | ||||||||||||||||||||||||||||||||||||||||||||
Db       6127 CTGTGCTTCTTCCTGCAGGTGAAGGTGGTGAGCCTGAAGCCTGAGGTGGCCCAGGTCGAC 6186

Qy        952 CTGTACATCCTCGGCCAAGCCGACCACTTTATTGGCAACTGTGTCTCCTCCTTCACTGCC 1011
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       6187 CTGTACATCCTCGGCCAAGCCGACCACTTTATTGGCAACTGTGTCTCCTCCTTCACTGCC 6246

Qy       1012 TTTGTGAAGCGGGAGCGGGACCTCCAGGGGAGGCCGTCTTCTTTCTTCGGCATGGACAGG 1071
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       6247 TTTGTGAAGCGGGAGCGGGACCTCCAGGGGAGGCCGTCTTCTTTCTTCGGCATGGACAGG 6306

Qy       1072 CCCCCTAAGCTGCGGGACGAGTTCTGATTCTGGCCGGAGCACCAGACCCTCTGATCCTGG 1131
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       6307 CCCCCTAAGCTGCGGGACGAGTTCTGATTCTGGCCGGAGCACCAGACCCTCTGATCCTGG 6366

Qy       1132 AGGGACCAGAGTCTGAGCTGGTCCTTCCAGCCAGGCCTGGCAGCCAGAGGTGCTCCGGGA 1191
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       6367 AGGGACCAGAGTCTGAGCTGGTCCTTCCAGCCAGGCCTGGCAGCCAGAGGTGCTCCGGGA 6426

Qy       1192 TTGCAAACTCCTCTTCTCACCTGCCAAAGATGGAGAAGAGTGCCAGGGACCCCTCAAGGA 1251
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       6427 TTGCAAACTCCTCTTCTCACCTGCCAAAGATGGAGAAGAGTGCCAGGGACCCCTCAAGGA 6486

Qy       1252 GGGAGACGCTCCATATCCCAGGGCATAGGACTTGCAGGTTCCTAGGAGCAGGAGCATCTC 1311
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       6487 GGGAGACGCTCCATATCCCAGGGCATAGGACTTGCAGGTTCCTAGGAGCAGGAGCATCTC 6546

Qy       1312 CCATCGCACGTGCTTTCTGCTCTTCTGGGAATTTCTCACACTGGCAAAGCAGTCCAGCCT 1371
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       6547 CCATCGCACGTGCTTTCTGCTCTTCTGGGAATTTCTCACACTGGCAAAGCAGTCCAGCCT 6606

Qy       1372 CCGTCTTCTGGTCCACTCTGCTCTGAGCAGCCTGGGATGCTGAACTCTTCAGAGAGATTT 1431
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       6607 CCGTCTTCTGGTCCACTCTGCTCTGAGCAGCCTGGGATGCTGAACTCTTCAGAGAGATTT 6666

Qy       1432 TTTTATAGAGAGATTTCTATAATTTTGATACAAGGTCATGACTATCCTAGAACTCTCTGT 1491
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       6667 TTTTATAGAGAGATTTCTATAATTTTGATACAAGGTCATGACTATCCTAGAACTCTCTGT 6726

Qy       1492 GGTTTTTGAAAATCATTGAATTC 1514
              |||||||||||||||||||||||
Db       6727 GGTTTTTGAAAATCATTGAATTC 6749


RESULT 11
ACH35532
ID   ACH35532 standard; cDNA; 477 BP.
XX
AC   ACH35532;
XX
DT   13-OCT-2003  (first entry)
XX
DE   Human endothelial cell cDNA #3665.
XX
```

```
KW   Human; ss; sequencing by hybridisation; SBH; expressed sequence tag; EST;
KW   genome mapping; biodiversity; genetic disorder.
XX
OS   Homo sapiens.
XX
PN   US2003073623-A1.
XX
PD   17-APR-2003.
XX
PF   30-JUL-2001; 2001US-00918995.
XX
PR   30-JUL-2001; 2001US-00918995.
XX
PA   (DRMA/) DRMANAC R T.
PA   (LABA/) LABAT I.
PA   (STAC/) STACHE-CRAIN B.
PA   (DICK/) DICKSON M C.
PA   (JONE/) JONES L W.
XX
PI   Drmanac RT,  Labat I,  Stache-Crain B,  Dickson MC,  Jones LW;
XX
DR   WPI; 2003-615964/58.
XX
PT   New polynucleotide sequences obtained from various cDNA libraries, useful
PT   as hybridization probes, as oligomers for PCR, for chromosome and gene
PT   mapping, in the recombinant production of protein, or in generating
PT   antisense DNA or RNA.
XX
PS   Claim 1; SEQ ID NO 22744; 44pp; English.
XX
CC   The invention relates to an isolated polynucleotide comprising any one of
CC   38043 cDNA sequences, appearing as ACH12789-ACH50831, whose sequence was
CC   determined by the technique of SBH (sequencing by hybridisation). Also
CC   included is a purified polypeptide comprising a sequence corresponding to
CC   a reading frame of the novel polynucleotide. The nucleic acid sequences
CC   are useful in diagnostics as expressed sequence tags (EST) for
CC   identifying expressed genes or for physical mapping of the human genome,
CC   in forensics, in assessing biodiversities, or in identifying mutations
CC   responsible for genetic disorders and other traits. The nucleotide
CC   sequences are also useful as hybridisation probes, as oligomers for PCR,
CC   for chromosome and gene mapping, in the recombinant production of
CC   protein, or in generating antisense DNA or RNA. The purified polypeptide
CC   is useful for generating antibodies specific for it. The present sequence
CC   is one of the 38043 isolated cDNA/EST sequences. Note: The sequence data
CC   for this patent did not form part of the printed specification, but was
CC   obtained in electronic format directly from USPTO at
CC   seqdata.uspto.gov/sequence.html?DocID=20030073623
XX
SQ   Sequence 477 BP; 95 A; 141 C; 142 G; 97 T; 0 U; 2 Other;

  Query Match             29.2%;  Score 442.8;  DB 9;  Length 477;
  Best Local Similarity   98.1%;  Pred. No. 2e-118;
  Matches  458;  Conservative    0;  Mismatches    8;  Indels    1;  Gaps    1;

Qy        855 TGCTACTGATTCCGAGAGTTATGTGCCTGAGCTCCAACAGCTCTTCAAAGGGAAGGTGAA 914
              || | ||||   | |||| ||||||||||||||| |||| ||||||||||||||||||
Db         11 TGATGCTGAGCNCTAGAGATATGTGCCTGAGCTCC-ACAGTTCTTCAAAGGGAAGGTGAA 69

Qy        915 GGTGGTGAGCCTGAAGCCTGAGGTGGCCCAGGTCGACCTGTACATCCTCGGCCAAGCCGA 974
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         70 GGTGGTGAGCCTGAAGCCTGAGGTGGCCCAGGTCGACCTGTACATCCTCGGCCAAGCCGA 129

Qy        975 CCACTTTATTGGCAACTGTGTCTCCTCCTTCACTGCCTTTGTGAAGCGGGAGCGGGACCT 1034
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        130 CCACTTTATTGGCAACTGTGTCTCCTCCTTCACTGCCTTTGTGAAGCGGGAGCGGGACCT 189

Qy       1035 CCAGGGGAGGCCGTCTTCTTTCTTCGGCATGGACAGGCCCCCTAAGCTGCGGGACGAGTT 1094
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        190 CCAGGGGAGGCCGTCTTCTTTCTTCGGCATGGACAGGCCCCCTAAGCTGCGGGACGAGTT 249

Qy       1095 CTGATTCTGGCCGGAGCACCAGACCCTCTGATCCTGGAGGGACCAGAGTCTGAGCTGGTC 1154
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        250 CTGATTCTGGCCGGAGCACCAGACCCTCTGATCCTGGAGGGACCAGAGTCTGAGCTGGTC 309

Qy       1155 CTTCCAGCCAGGCCTGGCAGCCAGAGGTGCTCCGGGATTGCAAACTCCTCTTCTCACCTG 1214
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        310 CTTCCAGCCAGGCCTGGCAGCCAGAGGTGCTCCGGGATTGCAAACTCCTCTTCTCACCTG 369

Qy       1215 CCAAAGATGGAGAAGAGTGCCAGGGACCCCTCAAGGAGGGAGACGCTCCATATCCCAGGG 1274
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        370 CCAAAGATGGAGAAGAGTGCCAGGGACCCCTCAAGGAGGGAGACGCTCCATATCCCAGGG 429

Qy       1275 CATAGGACTTGCAGGTTCCTAGGAGCAGGAGCATCTCCCATCGCACG 1321
              |||||||||||||||||||||||||||||||||||||||||||||
Db        430 CATAGGACTTGCAGGTTCCTAGGAGCAGGAGCATCTCCCATCGCACG 476


RESULT 12
ACH32225
ID   ACH32225 standard; cDNA; 479 BP.
XX
AC   ACH32225;
XX
DT   13-OCT-2003  (first entry)
XX
DE   Human endothelial cell cDNA #358.
XX
KW   Human; ss; sequencing by hybridisation; SBH; expressed sequence tag; EST;
KW   genome mapping; biodiversity; genetic disorder.
XX
OS   Homo sapiens.
XX
PN   US2003073623-A1.
XX
PD   17-APR-2003.
XX
PF   30-JUL-2001; 2001US-00918995.
XX
PR   30-JUL-2001; 2001US-00918995.
XX
PA   (DRMA/) DRMANAC R T.
PA   (LABA/) LABAT I.
PA   (STAC/) STACHE-CRAIN B.
PA   (DICK/) DICKSON M C.
PA   (JONE/) JONES L W.
XX
PI   Drmanac RT,  Labat I,  Stache-Crain B,  Dickson MC,  Jones LW;
XX
DR   WPI; 2003-615964/58.
XX
PT   New polynucleotide sequences obtained from various cDNA libraries, useful
PT   as hybridization probes, as oligomers for PCR, for chromosome and gene
PT   mapping, in the recombinant production of protein, or in generating
PT   antisense DNA or RNA.
XX
PS   Claim 1; SEQ ID NO 19437; 44pp; English.
XX
CC   The invention relates to an isolated polynucleotide comprising any one of
CC   38043 cDNA sequences, appearing as ACH12789-ACH50831, whose sequence was
CC   determined by the technique of SBH (sequencing by hybridisation). Also
CC   included is a purified polypeptide comprising a sequence corresponding to
CC   a reading frame of the novel polynucleotide. The nucleic acid sequences
CC   are useful in diagnostics as expressed sequence tags (EST) for
CC   identifying expressed genes or for physical mapping of the human genome,
CC   in forensics, in assessing biodiversities, or in identifying mutations
CC   responsible for genetic disorders and other traits. The nucleotide
CC   sequences are also useful as hybridisation probes, as oligomers for PCR,
CC   for chromosome and gene mapping, in the recombinant production of
CC   protein, or in generating antisense DNA or RNA. The purified polypeptide
CC   is useful for generating antibodies specific for it. The present sequence
CC   is one of the 38043 isolated cDNA/EST sequences. Note: The sequence data
CC   for this patent did not form part of the printed specification, but was
CC   obtained in electronic format directly from USPTO at
CC   seqdata.uspto.gov/sequence.html?DocID=20030073623
XX
SQ   Sequence 479 BP; 99 A; 140 C; 143 G; 97 T; 0 U; 0 Other;
```

```
  Query Match             29.1%;  Score 441;  DB 9;  Length 479;
  Best Local Similarity   98.9%;  Pred. No. 6.9e-118;
  Matches  444;  Conservative    0;  Mismatches    5;  Indels    0;  Gaps    0;

Qy        868 GAGAGTTATGTGCCTGAGCTCCAACAGCTCTTCAAAGGGAAGGTGAAGGTGGTGAGCCTG 927
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         31 GAGAGTTATGTGCCTGAGCTCCAACAGCTCTTCAAAGGGAAGGTGAAGGTGGTGAGCCTG 90

Qy        928 AAGCCTGAGGTGGCCCAGGTCGACCTGTACATCCTCGGCCAAGCCGACCACTTTATTGGC 987
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         91 AAGCCTGAGGTGGCCCAGGTCGACCTGTACATCCTCGGCCAAGCCGACCACTTTATTGGC 150

Qy        988 AACTGTGTCTCCTCCTTCACTGCCTTTGTGAAGCGGGAGCGGGACCTCCAGGGGAGGCCG 1047
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        151 AACTGTGTCTCCTCCTTCACTGCCTTTGTGAAGCGGGAGCGGGACCTCCAGGGGAGGCCG 210

Qy       1048 TCTTCTTTCTTCGGCATGGACAGGCCCCCTAAGCTGCGGGACGAGTTCTGATTCTGGCCG 1107
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        211 TCTTCTTTCTTCGGCATGGACAGGCCCCCTAAGCTGCGGGACGAGTTCTGATTCTGGCCG 270

Qy       1108 GAGCACCAGACCCTCTGATCCTGGAGGGACCAGAGTCTGAGCTGGTCCTTCCAGCCAGGC 1167
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        271 GAGCACCAGACCCTCTGATCCTGGAGGGACCAGAGTCTGAGCTGGTCCTTCCAGCCAGGC 330

Qy       1168 CTGGCAGCCAGAGGTGCTCCGGGATTGCAAACTCCTCTTCTCACCTGCCAAAGATGGAGA 1227
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        331 CTGGCAGCCAGAGGTGCTCCGGGATTGCAAACTCCTCTTCTCACCTGCCAAAGATGGAGA 390

Qy       1228 AGAGTGCCAGGGACCCCTCAAGGAGGGAGACGCTCCATATCCCAGGGCATAGGACTTGCA 1287
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||  |  |
Db        391 AGAGTGCCAGGGACCCCTCAAGGAGGGAGACGCTCCATATCCCAGGGCATAGGAACTTGA 450

Qy       1288 GGTTCCTAGGAGCAGGAGCATCTCCCATC 1316
              |||||||| ||||||||||||||||||||
Db        451 GGTTCCTAAGAGCAGGAGCATCTCCCATC 479
```

```
RESULT 13
ADQ56824
ID   ADQ56824 standard; DNA; 610 BP.
XX
AC   ADQ56824;
XX
DT   21-OCT-2004  (first entry)
XX
DE   Novel canine microarray-related DNA sequence SeqID8126.
XX
KW   canine microarray; drug screening; toxicity assay;
KW   environmental pollutant; cellular response; gene expression profile;
KW   toxic response; liver necrosis; fatty liver disease;
KW   protein adduct formation; hepatitis; dog; ds.
XX
OS   Canis familiaris.
XX
PN   WO2004063324-A2.
XX
PD   29-JUL-2004.
XX
PF   05-MAY-2003; 2003WO-US013853.
XX
PR   03-MAY-2002; 2002US-0377240P.
XX
PA   (GENE-) GENE LOGIC INC.
PA   (PFIZ ) PFIZER PROD INC.
XX
PI   Diggans JC,  Porter M,  Wei T;
XX
DR   WPI; 2004-561890/54.
XX
PT   New isolated nucleic acid molecule, useful for drug screening and
PT   toxicity assays or for assessing the impact, including toxicity, of a
PT   compound, pharmaceutical agent or environmental pollutant on a cell or
PT   living organism.
XX
PS   Claim 1; SEQ ID NO 8126; 41pp; English.
XX
CC   This invention is related to a novel isolated canine nucleic acid
CC   sequences and the construction of canine microarrays containing a
CC   significant portion of the canine genome. The isolated canine nucleic
CC   acid sequences of the invention may be useful for drug screening and
CC   toxicity assays. The invention is therefore useful for assessing the
CC   impact, including toxicity, of a compound, pharmaceutical agent or
CC   environmental pollutant on a cell or living organism. The methods are
CC   useful for detecting genes that are up- or down-regulated in canines in a
CC   disease state. The sequences are useful as diagnostic agents or markers
CC   to detect a cellular response in a sample individually or as part of a
CC   gene expression profile. It is also useful as a target for agents that
CC   modulate gene expression or activity. The database is useful for
CC   producing electronic Northerns that allow the user to determine the cell
CC   type or tissue in which a given gene is expressed and to allow
CC   determination of the abundance or expression level of a given gene in a
CC   particular tissue or cell. The methods are useful for determining the
CC   similarity of a toxic response to one or more individual compounds. The
CC   methods are useful for predicting at least one toxic response or the
CC   likelihood that a compound or test agent will induce various specific
CC   pathologies such as those of the liver (liver necrosis, fatty liver
CC   disease, protein adduct formation or hepatitis), those of the kidney,
CC   heart, brain or testes, or other pathologies associated with at least one
CC   of the toxins. The methods are also useful for predicting or elucidating
CC   the potential cellular pathways influenced, induced or modulated by the
CC   compound or test agent due to the similarity of the expression profile
CC   compared to the profile induced by a known toxin. The present sequence is
CC   that of a canine DNA sequence which was claimed for use during the
CC   production of a canine microarray of the invention.
XX
SQ   Sequence 610 BP; 155 A; 171 C; 149 G; 129 T; 0 U; 6 Other;
```

```
  Query Match             26.8%;  Score 406.2;  DB 13;  Length 610;
  Best Local Similarity   89.0%;  Pred. No. 1.1e-107;
  Matches  446;  Conservative    0;  Mismatches   54;  Indels    1;  Gaps    1;

Qy        171 CACCAACCTCCATGTGTCCTACCAGAAGTACTTCAAGCTGGAGCCCCTCCAGGCTTACCA 230
              |||||| |||||||||| ||||||||||||||||||||  |||||||| |||  |||||
Db          1 CACCAATCTCCATGTGTNCTACCAGAAGTACTTCAAGCTNNAGCCCCTCNAGGNNTACCA 60

Qy        231 TCGGGTCATCAGCTTGGAGGATTTCATGGAGAAGCTGGCACCCACCCACTGGCCCCCTGA 290
              |||||| |||||| | ||||| |||||||||||||||||||||| |||||||||| ||
Db         61 TCGGGTTATCAGCCTAGAGGACTTCATGGAGAAGCTGGCACCCACTCACTGGCCCCCCGA 120

Qy        291 GAAGCGGGTGGCATACTGCTTTGAGGTGGCAGCCCAGCGAAGCCCAGATAAGAAGACGTG 350
              || ||| |||||||||||||||| |||||||||||||||||||| |||||||||| ||
Db        121 GAGGCGAGTGGCATACTGCTTTGAAGTGGCAGCCCAGCGAAGCCCTGATAAGAAGACATG 180

Qy        351 CCCCATGAAGGAAGGAAACCCCTTTGGCCCATTCTGGGATCAGTTTCATGTGAGTTTCAA 410
              |||||||||||||||||| |||||||||||| ||||||||||||||||||||||||||||
Db        181 CCCCATGAAGGAAGGAAATCCCTTTGGCCCATTTTGGGATCAGTTTCATGTGAGTTTCAA 240

Qy        411 CAAGTCGGAGCTTTTTACAGGCATTTCCTTCAGTGCTTCCTACAGAGAACAATGGAGCCA 470
              ||||||  |||||||| |||||||||||||||||| ||||||| ||| |||||||  ||
Db        241 CAAGTCCAAGCTTTTTTCAGGCATTTCCTTCAGTGCCTCCTACAAAGACCAATGGATTCA 300

Qy        471 GAGATTTTCTCCAAAGGAACATCCGGTGCTTGCCCTGCCAGGAGCCCCAGCCCAGTTCCC 530
              |||||||||||||||||||||| ||||| |||| || |||||||| ||||||||||
Db        301 GAGATTTTCTCCAAAGGAACATCCAGTGCTGGCCC-CCCGGGAGCCCCCGCCCAGTTCCC 359

Qy        531 CGTCCTAGAAGAACACAGGCCACTACAGAAGTACATGGTATGGTCAGACGAAATGGTGAA 590
              |||||| || || ||||| ||||| |||||||| |||||||||||| || |||||||
Db        360 CGTCCTGGAGGAGCACAGACCACTCCAGAAGTATGTGGTATGGTCAGATGAGATGGTGAG 419

Qy        591 GACGGGAGAGGCCCAGATTCATGCCCACCTTGTCCGGCCCTATGTGGGCATTCATCTGCG 650
              |||||||||||||| |||| ||||||| ||||||| || |||||||||| |||||
Db        420 GACGGGAGAGGCCCACATTCTCACCCACCTCATCCGGCCGTACGTGGGCATTCACCTGCG 479

Qy        651 CATTGGCTCTGACTGGAAGAA 671
              |||||||||||||||| | |
```

```
Db        480 CATTGGCTCTGACTGGATGGA 500


RESULT 14
ABL13947
ID   ABL13947 standard; cDNA; 1209 BP.
XX
AC   ABL13947;
XX
DT   26-MAR-2002  (first entry)
XX
DE   Drosophila melanogaster expressed polynucleotide SEQ ID NO 36323.
XX
KW   Drosophila; developmental biology; cell signalling; insecticide;
KW   pharmaceutical; gene; ss.
XX
OS   Drosophila melanogaster.
XX
PN   WO200171042-A2.
XX
PD   27-SEP-2001.
XX
PF   23-MAR-2001; 2001WO-US009231.
XX
PR   23-MAR-2000; 2000US-0191637P.
PR   11-JUL-2000; 2000US-00614150.
XX
PA   (PEKE ) PE CORP NY.
XX
PI   Venter JC,  Adams M,  Li PWD,  Myers EW;
XX
DR   WPI; 2001-656860/75.
DR   P-PSDB; ABB69844.
XX
PT   New isolated nucleic acid detection reagent for detecting 1000 or more
PT   genes from Drosophila and for elucidating cell signaling and cell-cell
PT   interactions.
XX
PS   Claim 1; SEQ ID NO 36323; 21pp + Sequence Listing; English.
XX
CC   The invention relates to an isolated nucleic acid detection reagent
CC   capable of detecting 1000 or more genes from Drosophila. The invention is
CC   useful in developmental biology and in elucidating cell signalling and
CC   cell-cell interactions in higher eukaryotes for the development of
CC   insecticides, therapeutics and pharmaceutical drugs. The invention
CC   discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC   sequences (ABL01840-ABL16175) and the encoded proteins (ABB57737-
CC   ABB72072). The sequence data for this patent did not form part of the
CC   printed specification, but was obtained in electronic format directly
CC   from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX
SQ   Sequence 1209 BP; 302 A; 313 C; 307 G; 287 T; 0 U; 0 Other;

  Query Match             11.0%;  Score 166.4;  DB 4;  Length 1209;
  Best Local Similarity   51.5%;  Pred. No. 1.7e-37;
  Matches  563;  Conservative    0;  Mismatches  486;  Indels   45;  Gaps    6;

Qy         17 GGGACCCGGCCGGTTACCTGCTCTACTGCCCCTGCATGGGGCGCTTTGGGAACCAGGCCG 76
              | || ||    || |||||   |||||| || || ||||| |||||||| ||||||||||
Db         80 GCGATCCCAATGGCTACCTCACCTACTGTCCGTGTATGGGACGCTTTGGCAACCAGGCCG 139

Qy         77 ATCACTTCTTGGGCTCTCTGGCATTTGCAAAGCTGCTAAACCGTACCTTGGCTGTCCCTC 136
              | |||||| |||| ||  |||| || || |||  ||| || || ||| ||   | || |
Db        140 ACCACTTCCTGGGATCATTGGCCTTCGCCAAGGCGCTTAATCGCACCCTGATCCTGCCGC 199

Qy        137 CTTGGATTGAGTACCAGCATCACAAGCCTCCTTTCACCAACCTCCATGTGTCCTACCAGA 196
              | ||| | ||||      ||| | |  |     |     ||  || ||  | | | | |
Db        200 CGTGGGTGGAGT------ATCGTAGGGGTGAACTGCGATCCCGTCAGGTACCGTTCAACA 253

Qy        197 AGTACTTCAAGCTGGAGCCCCTCCAGGCTTACCATCGGGTCATCAGCTTGGAGGATTTCA 256
                || ||  |  |||||||||||  |||  |||||||| ||||||| | |||  |||||||
Db        254 CATATTTTGAAGTGGAGCCCCTGAAGGAATACCATCGCGTCATCACCATGGCAGATTTCA 313

Qy        257 TGGAGAAGCTGGCACCCACCCACTGGCCCCCTGAGAAGCGGGTGGCATACTGCTTTGAGG 316
              ||  | | |||||   |  |  |||||      | |||| ||| |||  ||||   |||
Db        314 TGTGGCACCTGGCCGACGACATTTGGCCAGAATCGGAGCGAGTGTCATTTTGCTACAAGG 373

Qy        317 --------------TGGCAGCCCAGCGAAGCCCAGATAAGAAGACGTGCCCCATGAAGG 361
                           |  |  ||     ||||| |||   |  |||| |   ||||
Db        374 AACGATATAGCCTTCAGCAGGAGAAGAACGATCCAGACAAGCCCAATTGCCACGCCAAGG 433

Qy        362 AAGGAAACCCCTTTGGCCCATTCTGGGATCAGTTTCATGTGAGTTTCAACAAGTCGGAGC 421
              | || || || ||||| || || |||||     |||||  |    ||        ||| |||
Db        434 ATGGCAATCCTTTTGGTCCCTTTTGGGACACTTTTCACATAGACTTTGTGCGGTCAGAGT 493

Qy        422 TTTTTACAGGCATTTCCTTCAGTGCTTCCTACAG---------AGAACAATGGAGCCAGA 472
              | | | |      ||   ||   ||       | ||                |  |||     |
Db        494 TCTATGCGCCACTTCATTTTGATGTGCATCATAGCAACGAGGCTGCCAAGTGGCAGACCA 553

Qy        473 GATTTTCTCCAAAGGAACATCCGGTGCTTGCCCTGCCAGGAGCCCCAGCCCAGTTCCCCG 532
                || | || || |   | |||| || || ||  |  | ||||| || ||    || || |
Db        554 AATATCCTGCAGAATCATATCCCGTACTCGCGTTCACCGGAGCTCCGGCTAGTTTTCCTG 613

Qy        533 TCCTAGAAGAACACAGGCCACTACAGAAGTACATGGTATGGTCAGACGAAATGGTGAAGA 592
              | |    |||  || |    || |||   ||| ||   |||    |        |  |
Db        614 TTCAGCTAGAGAACTGCAAGCTGCAGCGCTACTTGCAGTGGAGTCAACGGTATAGGGAAG 673

Qy        593 CGGGAGAGGCCCAGATTCATGCCCACCTTGTCCGGC---CCTATGTGGGCATTCATCTGC 649
              |      |||      || | |  || |    |||    ||| | ||||||||||||||||
Db        674 CATCTAAGGATTTCATCCGAGAGCAGTTGCCTCGGGGTGCCTTTTTGGGCATTCATCTGC 733

Qy        650 GCATTGGCTCTGACTGGAAGAACGCCTGTGCCATGCTGAAGGACGG---------GACTG 700
              |||  ||    || |||  ||  |||||||      | |||||  |         |  ||
Db        734 GCAACGGTATCGATTGGGTGAGAGCCTGTGAGCACGTCAAGGATAGCCAGCATCTGTTTG 793

Qy        701 CAGGCTCGCACTTCATGGCCTCTCCGCAGTGTGTGGGCTACAGCCGCAGCACAGCGGCCC 760
              |      |||| | | ||| || |   |        ||       |  |          |
Db        794 CCTCGCCGCAGTGCTTGGGCTATAAAAATGAACGTGGTGCACTCTACCCGGAGCTCTGCA 853

Qy        761 CCCTCACGATGACTATGTGCCTGCCTGACCTGAAGGAGATCCAGAGGGCTGTGAAGCTCT 820
              | | | | |     |  | | |   | || |||   | |   | |   |||   |
Db        854 TGCCCTCCAAGGAGGCGATCATCCGCCAGCTAAAGAGAACCATTAAGAACGTGCGCCAAA 913

Qy        821 GGGTGAGGTCGCTGGATGCCCAGTCGGTCTACGTTGCTACTGATTCCGAGAGTTATGTGC 880
                |  |        ||   | | || || | ||| ||  | || |||  |         |
Db        914 CTCAGCCGGACAACGAAATCAAATCAGTTTTCGTGGCGTCAGACTCCAATCACATGATTG 973

Qy        881 CTGAGCTCCAACAGCTCTTCAAAGGGAAGGTGAAGGTGGTGAGCCTGAAGCCTGAGGTGG 940
              ||| ||  |   |  | | |  | || |   |||  |  |  ||| ||||  |
Db        974 GTGAACTAAACACGGCCCTTAGTCGCATGGGCATCAGTGTGCACAAGCTGCCGGAGGATG 1033

Qy        941 ---CCCAGGTCGACCTGTACATCCTCGGCCAAGCCGACCACTTTATTGGCAACTGTGTCT 997
                |  |  | ||| ||  ||| ||||| ||  |  |||||||||| |||||||| | |
Db       1034 ATCCTTACCTGGACTTGGCCATTCTCGGACAGTCGAACCACTTTATCGGCAACTGTATAT 1093

Qy        998 CCTCCTTCACTGCCTTTGTGAAGCGGGAGCGGGACCTCCAGGGGAGGCCGTCTTCTTTCT 1057
              |||| | | || || |  ||  |||| || ||  | || ||    || || |  ||||
Db       1094 CCTCTTACTCGGCATTCGAAAAAAGGGAACGAGATGTGCACGGTTTTCCATCGTACTTCT 1153

Qy       1058 TCGGCATGGACAGG 1071
                ||  |   || |
Db       1154 GGGGATTCCCCAAG 1167


RESULT 15
ABL13946/c
ID   ABL13946 standard; cDNA; 3264 BP.
XX
AC   ABL13946;
XX
DT   26-MAR-2002  (first entry)
XX
DE   Drosophila melanogaster expressed polynucleotide SEQ ID NO 36320.
XX
KW   Drosophila; developmental biology; cell signalling; insecticide;
KW   pharmaceutical; gene; ss.
```

XX
OS Drosophila melanogaster.
XX
PN WO200171042-A2.
XX
PD 27-SEP-2001.
XX
PF 23-MAR-2001; 2001WO-US009231.
XX
PR 23-MAR-2000; 2000US-0191637P.
PR 11-JUL-2000; 2000US-00614150.
XX
PA (PEKE) PE CORP NY.
XX
PI Venter JC, Adams M, Li PWD, Myers EW;
XX
DR WPI; 2001-656860/75.
DR P-PSDB; ABB69843.
XX
PT New isolated nucleic acid detection reagent for detecting 1000 or more
PT genes from Drosophila and for elucidating cell signaling and cell-cell
PT interactions.
XX
PS Claim 1; SEQ ID NO 36320; 21pp + Sequence Listing; English.
XX
CC The invention relates to an isolated nucleic acid detection reagent
CC capable of detecting 1000 or more genes from Drosophila. The invention is
CC useful in developmental biology and in elucidating cell signalling and
CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics, and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC sequences (ABL01840-ABL16175) and the encoded proteins (ABB57737-
CC ABB72072). The sequence data for this patent did not form part of the
CC printed specification, but was obtained in electronic format directly
CC from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 3264 BP; 904 A; 742 C; 758 G; 860 T; 0 U; 0 Other;

Query Match 9.9%; Score 150.4; DB 4; Length 3264;
Best Local Similarity 50.9%; Pred. No. 1.4e-32;
Matches 541; Conservative 0; Mismatches 476; Indels 45; Gaps 6;

```
Qy        49 TGCATGGGGCGCTTTGGGAACCAGGCCGATCACTTCTTGGGCTCTCTGGCATTTGCAAAG 108
             | ||  || |||||||| |||||||||| |||||| |||| ||  |||| || || |||
Db      2098 TTCAAAGGACGCTTTGGCAACCAGGCCGACCACTTCCTGGGATCATTGGCCTTCGCCAAG 2039

Qy       109 CTGCTAAACCGTACCTTGGCTGTCCCTCCTTGGATTGAGTACCAGCATCACAAGCCTCCT 168
               ||| || || ||| ||    | || || ||| | ||||       |||  | |  |
Db      2038 GCGCTTAATCGCACCCTGATCCTGCCGCCGTGGGTGGAGT------ATCGTAGGGGTGAA 1985

Qy       169 TTCACCAACCTCCATGTGTCCTACCAGAAGTACTTCAAGCTGGAGCCCCTCCAGGCTTAC 228
               |     ||  || ||  | | | | |  || ||  |  ||||||||||  |||  |||
Db      1984 CTGCGATCCCGTCAGGTACCGTTCAACACATATTTTGAAGTGGAGCCCCTGAAGGAATAC 1925

Qy       229 CATCGGGTCATCAGCTTGGAGGATTTCATGGAGAAGCTGGCACCCACCCACTGGCCCCCT 288
             |||||  |||||||  | |||   |||||||||   | | |||||    |   |    |||||
Db      1924 CATCGCGTCATCACCATGGCAGATTTCATGTGGCACCTGGCCGACGACATTTGGCCAGAA 1865

Qy       289 GAGAAGCGGGTGGCATACTGCTTTGAGG----------------TGGCAGCCCAGCGAAGC 333
               | |||| ||| |||  ||||   |||                  |    |   ||
Db      1864 TCGGAGCGAGTGTCATTTTGCTACAAGGAACGATATAGCCTTCAGCAGGAGAAGAACGAT 1805

Qy       334 CCAGATAAGAAGACGTGCCCCATGAAGGAAGGAAACCCCCTTTGGCCCATTCTGGGATCAG 393
             |||||  |||    |  |||| |   ||||| || || || ||||| || || |||||
Db      1804 CCAGACAAGCCCAATTGCCACGCCAAGGATGGCAATCCTTTTGGTCCCTTTTGGGACACT 1745

Qy       394 TTTCATGTGAGTTTCAACAAGTCGGAGCTTTTTACAGGCATTTCCTTCAGTGCTTCCTAC 453
             |||||  |    ||      ||| ||| | | | |      ||   ||   ||       |
Db      1744 TTTCACATAGACTTTGTGCGGTCAGAGTTCTATGCGCCACTTCATTTTGATGTGCATCAT 1685

Qy       454 AG---------AGAACAATGGAGCCAGAGATTTTCTCCAAAGGAACATCCGGTGCTTGCC 504
             ||         |   | |||       | || | || || |   | |||| || || ||
Db      1684 AGCAACGAGGCTGCCAAGTGGCAGACCAAATATCCTGCAGAATCATATCCCGTACTCGCG 1625

Qy       505 CTGCCAGGAGCCCCAGCCCAGTTCCCCGTCCTAGAAGAACACAGGCCACTACAGAAGTAC 564
               | | ||||| || ||    || || || |    |||  || |    || |||   |||
Db      1624 TTCACCGGAGCTCCGGCTAGTTTTCCTGTTCAGCTAGAGAACTGCAAGCTGCAGCGCTAC 1565

Qy       565 ATGGTATGGTCAGACGAAATGGTGAAGACGGGAGAGGCCCAGATTCATGCCCACCTTGTC 624
             ||   |||    |        | |  |     |||     || |  |  ||  |
Db      1564 TTGCAGTGGAGTCAACGGTATAGGGAAGCATCTAAGGATTTCATCCGAGAGCAGTTGCCT 1505

Qy       625 CGG---CCCTATGTGGGCATTCATCTGCGCATTGGCTCTGACTGGAAGAACGCCTGTGCC 681
             |||    ||| | ||||||||||||||||||  ||     || |||  ||  ||||||||
Db      1504 CGGGGTGCCTTTTTGGGCATTCATCTGCGCAACGGTATCGATTGGGTGAGAGCCTGTGAG 1445

Qy       682 ATGCTGAAGGACGG---------GACTGCAGGCTCGCACTTCATGGCCTCTCCGCAGTGT 732
               | |||||  |         |  |||      |||| | | ||| || |    |
Db      1444 CACGTCAAGGATAGCCAGCATCTGTTTGCCTCGCCGCAGTGCTTGGGCTATAAAAATGAA 1385

Qy       733 GTGGGCTACAGCCGCAGCACAGCGGCCCCCCTCACGATGACTATGTGCCTGCCTGACCTG 792
               ||      |  |             |   | | | | |     |  | | |   | ||
Db      1384 CGTGGTGCACTCTACCCGGAGCTCTGCATGCCCTCCAAGGAGGCGATCATCCGCCAGCTA 1325

Qy       793 AAGGAGATCCAGAGGGCTGTGAAGCTCTGGGTGAGGTCGCTGGATGCCCAGTCGGTCTAC 852
             |||   | |    | |   |||   |        |  |       ||   | | || || | |
Db      1324 AAGAGAACCATTAAGAACGTGCGCCAAACTCAGCCGGACAACGAAATCAAATCAGTTTTC 1265

Qy       853 GTTGCTACTGATTCCGAGAGTTATGTGCCTGAGCTCCAACAGCTCTTCAAAGGGAAGGTG 912
             || ||  | || ||| |        |   ||| ||  |   |  | | |   | | ||
Db      1264 GTGGCGTCAGACTCCAATCACATGATTGGTGAACTAAACACGGCCCTTAGTCGCATGGGC 1205

Qy       913 AAGGTGGTGAGCCTGAAGCCTGAGGTGG---CCCAGGTCGACCTGTACATCCTCGGCCAA 969
             |     |||  |  |  ||| ||||  |   |  |  | ||| ||  ||| ||||| ||
Db      1204 ATCAGTGTGCACAAGCTGCCGGAGGATGATCCTTACCTGGACTTGGCCATTCTCGGACAG 1145

Qy       970 GCCGACCACTTTATTGGCAACTGTGTCTCCTCCTTCACTGCCTTTGTGAAGCGGGAGCGG 1029
               |  |||||||||| |||||||| | ||||| | | | || || |  ||  |||| ||
Db      1144 TCGAACCACTTTATCGGCAACTGTATATCCTCTTACTCGGCATTCGAAAAAAGGGAACGA 1085

Qy       1030 GACCTCCAGGGGAGGCCGTCTTCTTTCTTCGGCATGGACAGG 1071
              ||  | || ||   || || |  |||| || |   || |
Db       1084 GATGTGCACGGTTTTCCATCGTACTTCTGGGGATTCCCCAAG 1043
```

Search completed: November 9, 2005, 04:43:18
Job time : 887.842 secs

GenCore version 5.1.6

OM nucleic - nucleic search, using sw model

Run on: November 9, 2005, 02:15:54 ; Search time 273.956 Seconds
(without alignments)
9042.764 Million cell updates/sec

Title: US-09-774-954-1
Perfect score: 1514
Sequence: 1 ATGCCCGCGGGCTCCTGGGA.....TTTTGAAAATCATTGAATTC 1514

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1202784 seqs, 818138359 residues

Total number of hits satisfying chosen parameters: 2405568

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_NA:*
1: /cgn2_6/ptodata/1/ina/5A_COMB.seq:*
2: /cgn2_6/ptodata/1/ina/5B_COMB.seq:*
3: /cgn2_6/ptodata/1/ina/6A_COMB.seq:*
4: /cgn2_6/ptodata/1/ina/6B_COMB.seq:*
5: /cgn2_6/ptodata/1/ina/PCTUS_COMB.seq:*
6: /cgn2_6/ptodata/1/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	1514	100.0	1514	3	US-08-978-741-1	Sequence 1, Appli
	2	1514	100.0	1514	3	US-09-333-729A-2	Sequence 2, Appli
	3	1512.4	99.9	11284	3	US-08-978-741-5	Sequence 5, Appli
	4	1447.4	95.6	5009	3	US-08-978-741-7	Sequence 7, Appli
	5	1447.4	95.6	5009	3	US-09-333-729A-8	Sequence 8, Appli
	6	1163.4	76.8	1300	3	US-08-978-741-4	Sequence 4, Appli
	7	1163.4	76.8	1300	3	US-09-333-729A-6	Sequence 6, Appli
	8	1100	72.7	1100	3	US-08-978-741-16	Sequence 16, Appl
	9	1100	72.7	1100	3	US-09-333-729A-16	Sequence 16, Appl
	10	168	11.1	1320	4	US-09-270-767-14353	Sequence 14353, A
	11	62.2	4.1	722	4	US-09-270-767-30507	Sequence 30507, A
c	12	58.6	3.9	998	4	US-09-270-767-28905	Sequence 28905, A
c	13	58.6	3.9	2034	4	US-09-270-767-13022	Sequence 13022, A
c	14	56	3.7	19227	4	US-09-949-016-12127	Sequence 12127, A
c	15	56	3.7	19228	4	US-09-949-016-16285	Sequence 16285, A
c	16	41.2	2.7	114793	4	US-10-148-806-3	Sequence 3, Appli
	17	40.6	2.7	2084	4	US-09-902-540-4387	Sequence 4387, Ap
c	18	40.6	2.7	29272	4	US-09-902-540-1217	Sequence 1217, Ap
	19	40	2.6	40	3	US-08-978-741-10	Sequence 10, Appl
	20	40	2.6	40	3	US-08-978-741-12	Sequence 12, Appl
	21	40	2.6	40	3	US-09-333-729A-9	Sequence 9, Appli
	22	39	2.6	1131	4	US-09-252-991A-76	Sequence 76, Appl
	23	39	2.6	1353	4	US-09-252-991A-88	Sequence 88, Appl
	24	38.6	2.5	2358	4	US-09-919-039-190	Sequence 190, App
	25	37.8	2.5	2154	4	US-09-902-540-4557	Sequence 4557, Ap
	26	37.8	2.5	24905	4	US-09-902-540-1225	Sequence 1225, Ap
c	27	37.8	2.5	4403765	3	US-09-103-840A-2	Sequence 2, Appli
c	28	37.8	2.5	4411529	3	US-09-103-840A-1	Sequence 1, Appli
c	29	37.4	2.5	18037	4	US-09-949-016-13678	Sequence 13678, A
c	30	37.4	2.5	77536	4	US-09-410-551B-1	Sequence 1, Appli
c	31	37.4	2.5	77536	4	US-09-940-316B-1	Sequence 1, Appli
	32	37.2	2.5	933	4	US-09-252-991A-11943	Sequence 11943, A
c	33	37.2	2.5	1218	4	US-09-252-991A-12011	Sequence 12011, A
	34	37.2	2.5	1650	4	US-09-252-991A-11977	Sequence 11977, A
	35	37.2	2.5	102053	4	US-09-949-016-13025	Sequence 13025, A
	36	37	2.4	1843	1	US-07-918-023-1	Sequence 1, Appli
	37	37	2.4	5962	6	5386025-5	Patent No. 5386025
	38	37	2.4	5962	6	5386025-5	Patent No. 5386025
	39	37	2.4	5975	1	US-08-404-354B-1	Sequence 1, Appli
	40	37	2.4	5975	1	US-08-314-083B-1	Sequence 1, Appli
	41	37	2.4	5975	1	US-08-435-675B-1	Sequence 1, Appli
	42	37	2.4	5975	1	US-08-336-257A-3	Sequence 3, Appli
	43	37	2.4	5975	3	US-08-884-599-1	Sequence 1, Appli
	44	36.8	2.4	3591	4	US-10-237-551-191	Sequence 191, App
	45	36.8	2.4	3591	4	US-10-237-551-221	Sequence 221, App

ALIGNMENTS

```
RESULT 1
US-08-978-741-1
; Sequence 1, Application US/08978741
; Patent No. 6100076
;  GENERAL INFORMATION:
;    APPLICANT:  Yang Wang, Michael W. Spellman
;    TITLE OF INVENTION:  O-Fucosyltransferase
;    NUMBER OF SEQUENCES:  17
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Genentech, Inc.
;      STREET:  1 DNA Way
;      CITY:  South San Francisco
;      STATE:  California
;      COUNTRY:  USA
;      ZIP:  94080
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  3.5 inch, 1.44 Mb floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  WinPatin (Genentech)
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/978,741
;      FILING DATE:  26-No. 6100076-1997
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/792498
;      FILING DATE:  31
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Svoboda, Craig G.
;      REGISTRATION NUMBER:  39,044
;      REFERENCE/DOCKET NUMBER:  P1041P1
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  650/225-1489
;      TELEFAX:  650/952-9881
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1514 base pairs
;      TYPE:  Nucleic Acid
;      STRANDEDNESS:  Single
;      TOPOLOGY:  Linear
US-08-978-741-1

  Query Match             100.0%;  Score 1514;  DB 3;  Length 1514;
  Best Local Similarity   100.0%;  Pred. No. 0;
  Matches 1514;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 ATGCCCGCGGGCTCCTGGGACCCGGCCGGTTACCTGCTCTACTGCCCCTGCATGGGGCGC 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGCCCGCGGGCTCCTGGGACCCGGCCGGTTACCTGCTCTACTGCCCCTGCATGGGGCGC 60
```

```
Qy        61 TTTGGGAACCAGGCCGATCACTTCTTGGGCTCTCTGGCATTTGCAAAGCTGCTAAACCGT 120
Db        61 TTTGGGAACCAGGCCGATCACTTCTTGGGCTCTCTGGCATTTGCAAAGCTGCTAAACCGT 120

Qy       121 ACCTTGGCTGTCCCTCCTTGGATTGAGTACCAGCATCACAAGCCTCCTTTCACCAACCTC 180
Db       121 ACCTTGGCTGTCCCTCCTTGGATTGAGTACCAGCATCACAAGCCTCCTTTCACCAACCTC 180

Qy       181 CATGTGTCCTACCAGAAGTACTTCAAGCTGGAGCCCCTCCAGGCTTACCATCGGGTCATC 240
Db       181 CATGTGTCCTACCAGAAGTACTTCAAGCTGGAGCCCCTCCAGGCTTACCATCGGGTCATC 240

Qy       241 AGCTTGGAGGATTTCATGGAGAAGCTGGCACCCACCCACTGGCCCCCTGAGAAGCGGGTG 300
Db       241 AGCTTGGAGGATTTCATGGAGAAGCTGGCACCCACCCACTGGCCCCCTGAGAAGCGGGTG 300

Qy       301 GCATACTGCTTTGAGGTGGCAGCCCAGCGAAGCCCAGATAAGAAGACGTGCCCCATGAAG 360
Db       301 GCATACTGCTTTGAGGTGGCAGCCCAGCGAAGCCCAGATAAGAAGACGTGCCCCATGAAG 360

Qy       361 GAAGGAAACCCCTTTGGCCCATTCTGGGATCAGTTTCATGTGAGTTTCAACAAGTCGGAG 420
Db       361 GAAGGAAACCCCTTTGGCCCATTCTGGGATCAGTTTCATGTGAGTTTCAACAAGTCGGAG 420

Qy       421 CTTTTTACAGGCATTTCCTTCAGTGCTTCCTACAGAGAACAATGGAGCCAGAGATTTTCT 480
Db       421 CTTTTTACAGGCATTTCCTTCAGTGCTTCCTACAGAGAACAATGGAGCCAGAGATTTTCT 480

Qy       481 CCAAAGGAACATCCGGTGCTTGCCCTGCCAGGAGCCCCAGCCCAGTTCCCCGTCCTAGAA 540
Db       481 CCAAAGGAACATCCGGTGCTTGCCCTGCCAGGAGCCCCAGCCCAGTTCCCCGTCCTAGAA 540

Qy       541 GAACACAGGCCACTACAGAAGTACATGGTATGGTCAGACGAAATGGTGAAGACGGGAGAG 600
Db       541 GAACACAGGCCACTACAGAAGTACATGGTATGGTCAGACGAAATGGTGAAGACGGGAGAG 600

Qy       601 GCCCAGATTCATGCCCACCTTGTCCGGCCCTATGTGGGCATTCATCTGCGCATTGGCTCT 660
Db       601 GCCCAGATTCATGCCCACCTTGTCCGGCCCTATGTGGGCATTCATCTGCGCATTGGCTCT 660

Qy       661 GACTGGAAGAACGCCTGTGCCATGCTGAAGGACGGGACTGCAGGCTCGCACTTCATGGCC 720
Db       661 GACTGGAAGAACGCCTGTGCCATGCTGAAGGACGGGACTGCAGGCTCGCACTTCATGGCC 720

Qy       721 TCTCCGCAGTGTGTGGGCTACAGCCGCAGCACAGCGGCCCCCCTCACGATGACTATGTGC 780
Db       721 TCTCCGCAGTGTGTGGGCTACAGCCGCAGCACAGCGGCCCCCCTCACGATGACTATGTGC 780

Qy       781 CTGCCTGACCTGAAGGAGATCCAGAGGGCTGTGAAGCTCTGGGTGAGGTCGCTGGATGCC 840
Db       781 CTGCCTGACCTGAAGGAGATCCAGAGGGCTGTGAAGCTCTGGGTGAGGTCGCTGGATGCC 840

Qy       841 CAGTCGGTCTACGTTGCTACTGATTCCGAGAGTTATGTGCCTGAGCTCCAACAGCTCTTC 900
Db       841 CAGTCGGTCTACGTTGCTACTGATTCCGAGAGTTATGTGCCTGAGCTCCAACAGCTCTTC 900

Qy       901 AAAGGGAAGGTGAAGGTGGTGAGCCTGAAGCCTGAGGTGGCCCAGGTCGACCTGTACATC 960
Db       901 AAAGGGAAGGTGAAGGTGGTGAGCCTGAAGCCTGAGGTGGCCCAGGTCGACCTGTACATC 960

Qy       961 CTCGGCCAAGCCGACCACTTTATTGGCAACTGTGTCTCCTCCTTCACTGCCTTTGTGAAG 1020
Db       961 CTCGGCCAAGCCGACCACTTTATTGGCAACTGTGTCTCCTCCTTCACTGCCTTTGTGAAG 1020

Qy      1021 CGGGAGCGGGACCTCCAGGGGAGGCCGTCTTCTTTCTTCGGCATGGACAGGCCCCCTAAG 1080
Db      1021 CGGGAGCGGGACCTCCAGGGGAGGCCGTCTTCTTTCTTCGGCATGGACAGGCCCCCTAAG 1080

Qy      1081 CTGCGGGACGAGTTCTGATTCTGGCCGGAGCACCAGACCCTCTGATCCTGGAGGGACCAG 1140
Db      1081 CTGCGGGACGAGTTCTGATTCTGGCCGGAGCACCAGACCCTCTGATCCTGGAGGGACCAG 1140

Qy      1141 AGTCTGAGCTGGTCCTTCCAGCCAGGCCTGGCAGCCAGAGGTGCTCCGGGATTGCAAACT 1200
Db      1141 AGTCTGAGCTGGTCCTTCCAGCCAGGCCTGGCAGCCAGAGGTGCTCCGGGATTGCAAACT 1200

Qy      1201 CCTCTTCTCACCTGCCAAAGATGGAGAAGAGTGCCAGGGACCCCTCAAGGAGGGAGACGC 1260
Db      1201 CCTCTTCTCACCTGCCAAAGATGGAGAAGAGTGCCAGGGACCCCTCAAGGAGGGAGACGC 1260

Qy      1261 TCCATATCCCAGGGCATAGGACTTGCAGGTTCCTAGGAGCAGGAGCATCTCCCATCGCAC 1320
Db      1261 TCCATATCCCAGGGCATAGGACTTGCAGGTTCCTAGGAGCAGGAGCATCTCCCATCGCAC 1320

Qy      1321 GTGCTTTCTGCTCTTCTGGGAATTTCTCACACTGGCAAAGCAGTCCAGCCTCCGTCTTCT 1380
Db      1321 GTGCTTTCTGCTCTTCTGGGAATTTCTCACACTGGCAAAGCAGTCCAGCCTCCGTCTTCT 1380

Qy      1381 GGTCCACTCTGCTCTGAGCAGCCTGGGATGCTGAACTCTTCAGAGAGATTTTTTTATAGA 1440
Db      1381 GGTCCACTCTGCTCTGAGCAGCCTGGGATGCTGAACTCTTCAGAGAGATTTTTTTATAGA 1440

Qy      1441 GAGATTTCTATAATTTTGATACAAGGTCATGACTATCCTAGAACTCTCTGTGGTTTTTGA 1500
Db      1441 GAGATTTCTATAATTTTGATACAAGGTCATGACTATCCTAGAACTCTCTGTGGTTTTTGA 1500

Qy      1501 AAATCATTGAATTC 1514
Db      1501 AAATCATTGAATTC 1514
```

RESULT 2
US-09-333-729A-2
; Sequence 2, Application US/09333729A
; Patent No. 6270987
; GENERAL INFORMATION:
; APPLICANT: Wang,Yang
; APPLICANT: Spellman,Michael W.
; TITLE OF INVENTION: O-Fucosyltransferase
; FILE REFERENCE: P1041P1D1-Substitute
; CURRENT APPLICATION NUMBER: US/09/333,729A
; CURRENT FILING DATE: 1999-06-15
; PRIOR APPLICATION NUMBER: US 08/798,741
; PRIOR FILING DATE: 1997-11-26
; NUMBER OF SEQ ID NOS: 21
; SEQ ID NO 2
; LENGTH: 1514
; TYPE: DNA
; ORGANISM: Homo Sapien
US-09-333-729A-2

Query Match 100.0%; Score 1514; DB 3; Length 1514;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1514; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 ATGCCCGCGGGCTCCTGGGACCCGGCCGGTTACCTGCTCTACTGCCCCTGCATGGGGCGC 60
Db          1 ATGCCCGCGGGCTCCTGGGACCCGGCCGGTTACCTGCTCTACTGCCCCTGCATGGGGCGC 60

Qy         61 TTTGGGAACCAGGCCGATCACTTCTTGGGCTCTCTGGCATTTGCAAAGCTGCTAAACCGT 120
Db         61 TTTGGGAACCAGGCCGATCACTTCTTGGGCTCTCTGGCATTTGCAAAGCTGCTAAACCGT 120

Qy        121 ACCTTGGCTGTCCCTCCTTGGATTGAGTACCAGCATCACAAGCCTCCTTTCACCAACCTC 180
Db        121 ACCTTGGCTGTCCCTCCTTGGATTGAGTACCAGCATCACAAGCCTCCTTTCACCAACCTC 180

Qy        181 CATGTGTCCTACCAGAAGTACTTCAAGCTGGAGCCCCTCCAGGCTTACCATCGGGTCATC 240
Db        181 CATGTGTCCTACCAGAAGTACTTCAAGCTGGAGCCCCTCCAGGCTTACCATCGGGTCATC 240

Qy        241 AGCTTGGAGGATTTCATGGAGAAGCTGGCACCCACCCACTGGCCCCCTGAGAAGCGGGTG 300
Db        241 AGCTTGGAGGATTTCATGGAGAAGCTGGCACCCACCCACTGGCCCCCTGAGAAGCGGGTG 300

Qy        301 GCATACTGCTTTGAGGTGGCAGCCCAGCGAAGCCCAGATAAGAAGACGTGCCCCATGAAG 360
```

```
Db        301 GCATACTGCTTTGAGGTGGCAGCCCAGCGAAGCCCAGATAAGAAGACGTGCCCCATGAAG 360

Qy        361 GAAGGAAACCCCTTTGGCCCATTCTGGGATCAGTTTCATGTGAGTTTCAACAAGTCGGAG 420

Db        361 GAAGGAAACCCCTTTGGCCCATTCTGGGATCAGTTTCATGTGAGTTTCAACAAGTCGGAG 420

Qy        421 CTTTTTACAGGCATTTCCTTCAGTGCTTCCTACAGAGAACAATGGAGCCAGAGATTTTCT 480

Db        421 CTTTTTACAGGCATTTCCTTCAGTGCTTCCTACAGAGAACAATGGAGCCAGAGATTTTCT 480

Qy        481 CCAAAGGAACATCCGGTGCTTGCCCTGCCAGGAGCCCCAGCCCAGTTCCCCGTCCTAGAA 540

Db        481 CCAAAGGAACATCCGGTGCTTGCCCTGCCAGGAGCCCCAGCCCAGTTCCCCGTCCTAGAA 540

Qy        541 GAACACAGGCCACTACAGAAGTACATGGTATGGTCAGACGAAATGGTGAAGACGGGAGAG 600

Db        541 GAACACAGGCCACTACAGAAGTACATGGTATGGTCAGACGAAATGGTGAAGACGGGAGAG 600

Qy        601 GCCCAGATTCATGCCCACCTTGTCCGGCCCTATGTGGGCATTCATCTGCGCATTGGCTCT 660

Db        601 GCCCAGATTCATGCCCACCTTGTCCGGCCCTATGTGGGCATTCATCTGCGCATTGGCTCT 660

Qy        661 GACTGGAAGAACGCCTGTGCCATGCTGAAGGACGGGACTGCAGGCTCGCACTTCATGGCC 720

Db        661 GACTGGAAGAACGCCTGTGCCATGCTGAAGGACGGGACTGCAGGCTCGCACTTCATGGCC 720

Qy        721 TCTCCGCAGTGTGTGGGCTACAGCCGCAGCACAGCGGCCCCCCTCACGATGACTATGTGC 780

Db        721 TCTCCGCAGTGTGTGGGCTACAGCCGCAGCACAGCGGCCCCCCTCACGATGACTATGTGC 780

Qy        781 CTGCCTGACCTGAAGGAGATCCAGAGGGCTGTGAAGCTCTGGGTGAGGTCGCTGGATGCC 840

Db        781 CTGCCTGACCTGAAGGAGATCCAGAGGGCTGTGAAGCTCTGGGTGAGGTCGCTGGATGCC 840

Qy        841 CAGTCGGTCTACGTTGCTACTGATTCCGAGAGTTATGTGCCTGAGCTCCAACAGCTCTTC 900

Db        841 CAGTCGGTCTACGTTGCTACTGATTCCGAGAGTTATGTGCCTGAGCTCCAACAGCTCTTC 900

Qy        901 AAAGGGAAGGTGAAGGTGGTGAGCCTGAAGCCTGAGGTGGCCCAGGTCGACCTGTACATC 960

Db        901 AAAGGGAAGGTGAAGGTGGTGAGCCTGAAGCCTGAGGTGGCCCAGGTCGACCTGTACATC 960

Qy        961 CTCGGCCAAGCCGACCACTTTATTGGCAACTGTGTCTCCTCCTTCACTGCCTTTGTGAAG 1020

Db        961 CTCGGCCAAGCCGACCACTTTATTGGCAACTGTGTCTCCTCCTTCACTGCCTTTGTGAAG 1020

Qy       1021 CGGGAGCGGGACCTCCAGGGGAGGCCGTCTTCTTTCTTCGGCATGGACAGGCCCCCTAAG 1080

Db       1021 CGGGAGCGGGACCTCCAGGGGAGGCCGTCTTCTTTCTTCGGCATGGACAGGCCCCCTAAG 1080

Qy       1081 CTGCGGGACGAGTTCTGATTCTGGCCGGAGCACCAGACCCTCTGATCCTGGAGGGACCAG 1140

Db       1081 CTGCGGGACGAGTTCTGATTCTGGCCGGAGCACCAGACCCTCTGATCCTGGAGGGACCAG 1140

Qy       1141 AGTCTGAGCTGGTCCTTCCAGCCAGGCCTGGCAGCCAGAGGTGCTCCGGGATTGCAAACT 1200

Db       1141 AGTCTGAGCTGGTCCTTCCAGCCAGGCCTGGCAGCCAGAGGTGCTCCGGGATTGCAAACT 1200

Qy       1201 CCTCTTCTCACCTGCCAAAGATGGAGAAGAGTGCCAGGGACCCCTCAAGGAGGGAGACGC 1260

Db       1201 CCTCTTCTCACCTGCCAAAGATGGAGAAGAGTGCCAGGGACCCCTCAAGGAGGGAGACGC 1260

Qy       1261 TCCATATCCCAGGGCATAGGACTTGCAGGTTCCTAGGAGCAGGAGCATCTCCCATCGCAC 1320

Db       1261 TCCATATCCCAGGGCATAGGACTTGCAGGTTCCTAGGAGCAGGAGCATCTCCCATCGCAC 1320

Qy       1321 GTGCTTTCTGCTCTTCTGGGAATTTCTCACACTGGCAAAGCAGTCCAGCCTCCGTCTTCT 1380

Db       1321 GTGCTTTCTGCTCTTCTGGGAATTTCTCACACTGGCAAAGCAGTCCAGCCTCCGTCTTCT 1380

Qy       1381 GGTCCACTCTGCTCTGAGCAGCCTGGGATGCTGAACTCTTCAGAGAGATTTTTTTATAGA 1440

Db       1381 GGTCCACTCTGCTCTGAGCAGCCTGGGATGCTGAACTCTTCAGAGAGATTTTTTTATAGA 1440

Qy       1441 GAGATTTCTATAATTTTGATACAAGGTCATGACTATCCTAGAACTCTCTGTGGTTTTTGA 1500

Db       1441 GAGATTTCTATAATTTTGATACAAGGTCATGACTATCCTAGAACTCTCTGTGGTTTTTGA 1500

Qy       1501 AAATCATTGAATTC 1514

Db       1501 AAATCATTGAATTC 1514
```

RESULT 3
US-08-978-741-5
; Sequence 5, Application US/08978741
; Patent No. 6100076
; GENERAL INFORMATION:
; APPLICANT: Yang Wang, Michael W. Spellman
; TITLE OF INVENTION: O-Fucosyltransferase
; NUMBER OF SEQUENCES: 17
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genentech, Inc.
; STREET: 1 DNA Way
; CITY: South San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94080
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5 inch, 1.44 Mb floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: WinPatin (Genentech)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/978,741
; FILING DATE: 26-No. 6100076-1997
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/792498
; FILING DATE: 31
; ATTORNEY/AGENT INFORMATION:
; NAME: Svoboda, Craig G.
; REGISTRATION NUMBER: 39,044
; REFERENCE/DOCKET NUMBER: P1041P1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 650/225-1489
; TELEFAX: 650/952-9881
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 11284 base pairs
; TYPE: Nucleic Acid
; STRANDEDNESS: Single
; TOPOLOGY: Linear
US-08-978-741-5

Query Match 99.9%; Score 1512.4; DB 3; Length 11284;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 1513; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```
Qy          1 ATGCCCGCGGGCTCCTGGGACCCGGCCGGTTACCTGCTCTACTGCCCCTGCATGGGGCGC 60

Db       4236 ATGCCCGCGGGCTCCTGGGACCCGGCCGGTTACCTGCTCTACTGCCCCTGCATGGGGCGC 4295

Qy         61 TTTGGGAACCAGGCCGATCACTTCTTGGGCTCTCTGGCATTTGCAAAGCTGCTAAACCGT 120

Db       4296 TTTGGGAACCAGGCCGATCACTTCTTGGGCTCTCTGGCATTTGCAAAGCTGCTAAACCGT 4355

Qy        121 ACCTTGGCTGTCCCTCCTTGGATTGAGTACCAGCATCACAAGCCTCCTTTCACCAACCTC 180

Db       4356 ACCTTGGCTGTCCCTCCTTGGATTGAGTACCAGCATCACAAGCCTCCTTTCACCAACCTC 4415

Qy        181 CATGTGTCCTACCAGAAGTACTTCAAGCTGGAGCCCCTCCAGGCTTACCATCGGGTCATC 240

Db       4416 CATGTGTCCTACCAGAAGTACTTCAAGCTGGAGCCCCTCCAGGCTTACCATCGGGTCATC 4475
```

```
Qy       241 AGCTTGGAGGATTTCATGGAGAAGCTGGCACCCACCCACTGGCCCCCTGAGAAGCGGGTG 300
Db      4476 AGCTTGGAGGATTTCATGGAGAAGCTGGCACCCACCCACTGGCCCCCTGAGAAGCGGGTG 4535

Qy       301 GCATACTGCTTTGAGGTGGCAGCCCAGCGAAGCCCAGATAAGAAGACGTGCCCCATGAAG 360
Db      4536 GCATACTGCTTTGAGGTGGCAGCCCAGCGAAGCCCAGATAAGAAGACGTGCCCCATGAAG 4595

Qy       361 GAAGGAAACCCCTTTGGCCCATTCTGGGATCAGTTTCATGTGAGTTTCAACAAGTCGGAG 420
Db      4596 GAAGGAAACCCCTTTGGCCCATTCTGGGATCAGTTTCATGTGAGTTTCAACAAGTCGGAG 4655

Qy       421 CTTTTTACAGGCATTTCCTTCAGTGCTTCCTACAGAGAACAATGGAGCCAGAGATTTTCT 480
Db      4656 CTTTTTACAGGCATTTCCTTCAGTGCTTCCTACAGAGAACAATGGAGCCAGAGATTTTCT 4715

Qy       481 CCAAAGGAACATCCGGTGCTTGCCCTGCCAGGAGCCCCAGCCCAGTTCCCCGTCCTAGAA 540
Db      4716 CCAAAGGAACATCCGGTGCTTGCCCTGCCAGGAGCCCCAGCCCAGTTCCCCGTCCTAGAG 4775

Qy       541 GAACACAGGCCACTACAGAAGTACATGGTATGGTCAGACGAAATGGTGAAGACGGGAGAG 600
Db      4776 GAACACAGGCCACTACAGAAGTACATGGTATGGTCAGACGAAATGGTGAAGACGGGAGAG 4835

Qy       601 GCCCAGATTCATGCCCACCTTGTCCGGCCCTATGTGGGCATTCATCTGCGCATTGGCTCT 660
Db      4836 GCCCAGATTCATGCCCACCTTGTCCGGCCCTATGTGGGCATTCATCTGCGCATTGGCTCT 4895

Qy       661 GACTGGAAGAACGCCTGTGCCATGCTGAAGGACGGGACTGCAGGCTCGCACTTCATGGCC 720
Db      4896 GACTGGAAGAACGCCTGTGCCATGCTGAAGGACGGGACTGCAGGCTCGCACTTCATGGCC 4955

Qy       721 TCTCCGCAGTGTGTGGGCTACAGCCGCAGCACAGCGGCCCCCCTCACGATGACTATGTGC 780
Db      4956 TCTCCGCAGTGTGTGGGCTACAGCCGCAGCACAGCGGCCCCCCTCACGATGACTATGTGC 5015

Qy       781 CTGCCTGACCTGAAGGAGATCCAGAGGGCTGTGAAGCTCTGGGTGAGGTCGCTGGATGCC 840
Db      5016 CTGCCTGACCTGAAGGAGATCCAGAGGGCTGTGAAGCTCTGGGTGAGGTCGCTGGATGCC 5075

Qy       841 CAGTCGGTCTACGTTGCTACTGATTCCGAGAGTTATGTGCCTGAGCTCCAACAGCTCTTC 900
Db      5076 CAGTCGGTCTACGTTGCTACTGATTCCGAGAGTTATGTGCCTGAGCTCCAACAGCTCTTC 5135

Qy       901 AAAGGGAAGGTGAAGGTGGTGAGCCTGAAGCCTGAGGTGGCCCAGGTCGACCTGTACATC 960
Db      5136 AAAGGGAAGGTGAAGGTGGTGAGCCTGAAGCCTGAGGTGGCCCAGGTCGACCTGTACATC 5195

Qy       961 CTCGGCCAAGCCGACCACTTTATTGGCAACTGTGTCTCCTCCTTCACTGCCTTTGTGAAG 1020
Db      5196 CTCGGCCAAGCCGACCACTTTATTGGCAACTGTGTCTCCTCCTTCACTGCCTTTGTGAAG 5255

Qy      1021 CGGGAGCGGGACCTCCAGGGGAGGCCGTCTTCTTTCTTCGGCATGGACAGGCCCCCTAAG 1080
Db      5256 CGGGAGCGGGACCTCCAGGGGAGGCCGTCTTCTTTCTTCGGCATGGACAGGCCCCCTAAG 5315

Qy      1081 CTGCGGGACGAGTTCTGATTCTGGCCGGAGCACCAGACCCTCTGATCCTGGAGGGACCAG 1140
Db      5316 CTGCGGGACGAGTTCTGATTCTGGCCGGAGCACCAGACCCTCTGATCCTGGAGGGACCAG 5375

Qy      1141 AGTCTGAGCTGGTCCTTCCAGCCAGGCCTGGCAGCCAGAGGTGCTCCGGGATTGCAAACT 1200
Db      5376 AGTCTGAGCTGGTCCTTCCAGCCAGGCCTGGCAGCCAGAGGTGCTCCGGGATTGCAAACT 5435

Qy      1201 CCTCTTCTCACCTGCCAAAGATGGAGAAGAGTGCCAGGGACCCCTCAAGGAGGGAGACGC 1260
Db      5436 CCTCTTCTCACCTGCCAAAGATGGAGAAGAGTGCCAGGGACCCCTCAAGGAGGGAGACGC 5495

Qy      1261 TCCATATCCCAGGGCATAGGACTTGCAGGTTCCTAGGAGCAGGAGCATCTCCCATCGCAC 1320
Db      5496 TCCATATCCCAGGGCATAGGACTTGCAGGTTCCTAGGAGCAGGAGCATCTCCCATCGCAC 5555

Qy      1321 GTGCTTTCTGCTCTTCTGGGAATTTCTCACACTGGCAAAGCAGTCCAGCCTCCGTCTTCT 1380
Db      5556 GTGCTTTCTGCTCTTCTGGGAATTTCTCACACTGGCAAAGCAGTCCAGCCTCCGTCTTCT 5615

Qy      1381 GGTCCACTCTGCTCTGAGCAGCCTGGGATGCTGAACTCTTCAGAGAGATTTTTTTATAGA 1440
Db      5616 GGTCCACTCTGCTCTGAGCAGCCTGGGATGCTGAACTCTTCAGAGAGATTTTTTTATAGA 5675

Qy      1441 GAGATTTCTATAATTTTGATACAAGGTCATGACTATCCTAGAACTCTCTGTGGTTTTTGA 1500
Db      5676 GAGATTTCTATAATTTTGATACAAGGTCATGACTATCCTAGAACTCTCTGTGGTTTTTGA 5735

Qy      1501 AAATCATTGAATTC 1514
Db      5736 AAATCATTGAATTC 5749
```

```
RESULT 4
US-08-978-741-7
; Sequence 7, Application US/08978741
; Patent No. 6100076
;  GENERAL INFORMATION:
;    APPLICANT:  Yang Wang, Michael W. Spellman
;    TITLE OF INVENTION:  O-Fucosyltransferase
;    NUMBER OF SEQUENCES:  17
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Genentech, Inc.
;      STREET:  1 DNA Way
;      CITY:  South San Francisco
;      STATE:  California
;      COUNTRY:  USA
;      ZIP:  94080
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  3.5 inch, 1.44 Mb floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  WinPatin (Genentech)
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/978,741
;      FILING DATE:  26-No. 6100076-1997
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/792498
;      FILING DATE:  31
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Svoboda, Craig G.
;      REGISTRATION NUMBER:  39,044
;      REFERENCE/DOCKET NUMBER:  P1041P1
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  650/225-1489
;      TELEFAX:  650/952-9881
;  INFORMATION FOR SEQ ID NO:  7:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  5009 base pairs
;      TYPE:  Nucleic Acid
;      STRANDEDNESS:  Single
;      TOPOLOGY:  Linear
US-08-978-741-7

  Query Match             95.6%;  Score 1447.4;  DB 3;  Length 5009;
  Best Local Similarity   99.9%;  Pred. No. 0;
  Matches 1448;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy         66 GAACCAGGCCGATCACTTCTTGGGCTCTCTGGCATTTGCAAAGCTGCTAAACCGTACCTT 125
Db          1 GAACCAGGCCGATCACTTCTTGGGCTCTCTGGCATTTGCAAAGCTGCTAAACCGTACCTT 60

Qy        126 GGCTGTCCCTCCTTGGATTGAGTACCAGCATCACAAGCCTCCTTTCACCAACCTCCATGT 185
Db         61 GGCTGTCCCTCCTTGGATTGAGTACCAGCATCACAAGCCTCCTTTCACCAACCTCCATGT 120

Qy        186 GTCCTACCAGAAGTACTTCAAGCTGGAGCCCCTCCAGGCTTACCATCGGGTCATCAGCTT 245
Db        121 GTCCTACCAGAAGTACTTCAAGCTGGAGCCCCTCCAGGCTTACCATCGGGTCATCAGCTT 180
```

```
Qy       246 GGAGGATTTCATGGAGAAGCTGGCACCCACCCACTGGCCCCCTGAGAAGCGGGTGGCATA 305
Db       181 GGAGGATTTCATGGAGAAGCTGGCACCCACCCACTGGCCCCCTGAGAAGCGGGTGGCATA 240

Qy       306 CTGCTTTGAGGTGGCAGCCCAGCGAAGCCCAGATAAGAAGACGTGCCCCATGAAGGAAGG 365
Db       241 CTGCTTTGAGGTGGCAGCCCAGCGAAGCCCAGATAAGAAGACGTGCCCCATGAAGGAAGG 300

Qy       366 AAACCCCTTTGGCCCATTCTGGGATCAGTTTCATGTGAGTTTCAACAAGTCGGAGCTTTT 425
Db       301 AAACCCCTTTGGCCCATTCTGGGATCAGTTTCATGTGAGTTTCAACAAGTCGGAGCTTTT 360

Qy       426 TACAGGCATTTCCTTCAGTGCTTCCTACAGAGAACAATGGAGCCAGAGATTTTCTCCAAA 485
Db       361 TACAGGCATTTCCTTCAGTGCTTCCTACAGAGAACAATGGAGCCAGAGATTTTCTCCAAA 420

Qy       486 GGAACATCCGGTGCTTGCCCTGCCAGGAGCCCCAGCCCAGTTCCCCGTCCTAGAAGAACA 545
Db       421 GGAACATCCGGTGCTTGCCCTGCCAGGAGCCCCAGCCCAGTTCCCCGTCCTAGAGGAACA 480

Qy       546 CAGGCCACTACAGAAGTACATGGTATGGTCAGACGAAATGGTGAAGACGGGAGAGGCCCA 605
Db       481 CAGGCCACTACAGAAGTACATGGTATGGTCAGACGAAATGGTGAAGACGGGAGAGGCCCA 540

Qy       606 GATTCATGCCCACCTTGTCCGGCCCTATGTGGGCATTCATCTGCGCATTGGCTCTGACTG 665
Db       541 GATTCATGCCCACCTTGTCCGGCCCTATGTGGGCATTCATCTGCGCATTGGCTCTGACTG 600

Qy       666 GAAGAACGCCTGTGCCATGCTGAAGGACGGGACTGCAGGCTCGCACTTCATGGCCTCTCC 725
Db       601 GAAGAACGCCTGTGCCATGCTGAAGGACGGGACTGCAGGCTCGCACTTCATGGCCTCTCC 660

Qy       726 GCAGTGTGTGGGCTACAGCCGCAGCACAGCGGCCCCCCTCACGATGACTATGTGCCTGCC 785
Db       661 GCAGTGTGTGGGCTACAGCCGCAGCACAGCGGCCCCCCTCACGATGACTATGTGCCTGCC 720

Qy       786 TGACCTGAAGGAGATCCAGAGGGCTGTGAAGCTCTGGGTGAGGTCGCTGGATGCCCAGTC 845
Db       721 TGACCTGAAGGAGATCCAGAGGGCTGTGAAGCTCTGGGTGAGGTCGCTGGATGCCCAGTC 780

Qy       846 GGTCTACGTTGCTACTGATTCCGAGAGTTATGTGCCTGAGCTCCAACAGCTCTTCAAAGG 905
Db       781 GGTCTACGTTGCTACTGATTCCGAGAGTTATGTGCCTGAGCTCCAACAGCTCTTCAAAGG 840

Qy       906 GAAGGTGAAGGTGGTGAGCCTGAAGCCTGAGGTGGCCCAGGTCGACCTGTACATCCTCGG 965
Db       841 GAAGGTGAAGGTGGTGAGCCTGAAGCCTGAGGTGGCCCAGGTCGACCTGTACATCCTCGG 900

Qy       966 CCAAGCCGACCACTTTATTGGCAACTGTGTCTCCTCCTTCACTGCCTTTGTGAAGCGGGA 1025
Db       901 CCAAGCCGACCACTTTATTGGCAACTGTGTCTCCTCCTTCACTGCCTTTGTGAAGCGGGA 960

Qy      1026 GCGGGACCTCCAGGGGAGGCCGTCTTCTTTCTTCGGCATGGACAGGCCCCCTAAGCTGCG 1085
Db       961 GCGGGACCTCCAGGGGAGGCCGTCTTCTTTCTTCGGCATGGACAGGCCCCCTAAGCTGCG 1020

Qy      1086 GGACGAGTTCTGATTCTGGCCGGAGCACCAGACCCTCTGATCCTGGAGGGACCAGAGTCT 1145
Db       1021 GGACGAGTTCTGATTCTGGCCGGAGCACCAGACCCTCTGATCCTGGAGGGACCAGAGTCT 1080

Qy      1146 GAGCTGGTCCTTCCAGCCAGGCCTGGCAGCCAGAGGTGCTCCGGGATTGCAAACTCCTCT 1205
Db       1081 GAGCTGGTCCTTCCAGCCAGGCCTGGCAGCCAGAGGTGCTCCGGGATTGCAAACTCCTCT 1140

Qy      1206 TCTCACCTGCCAAAGATGGAGAAGAGTGCCAGGGACCCCTCAAGGAGGGAGACGCTCCAT 1265
Db       1141 TCTCACCTGCCAAAGATGGAGAAGAGTGCCAGGGACCCCTCAAGGAGGGAGACGCTCCAT 1200

Qy      1266 ATCCCAGGGCATAGGACTTGCAGGTTCCTAGGAGCAGGAGCATCTCCCATCGCACGTGCT 1325
Db       1201 ATCCCAGGGCATAGGACTTGCAGGTTCCTAGGAGCAGGAGCATCTCCCATCGCACGTGCT 1260

Qy      1326 TTCTGCTCTTCTGGGAATTTCTCACACTGGCAAAGCAGTCCAGCCTCCGTCTTCTGGTCC 1385
Db       1261 TTCTGCTCTTCTGGGAATTTCTCACACTGGCAAAGCAGTCCAGCCTCCGTCTTCTGGTCC 1320

Qy      1386 ACTCTGCTCTGAGCAGCCTGGGATGCTGAACTCTTCAGAGAGATTTTTTTATAGAGAGAT 1445
Db       1321 ACTCTGCTCTGAGCAGCCTGGGATGCTGAACTCTTCAGAGAGATTTTTTTATAGAGAGAT 1380

Qy      1446 TTCTATAATTTTGATACAAGGTCATGACTATCCTAGAACTCTCTGTGGTTTTTGAAAATC 1505
Db       1381 TTCTATAATTTTGATACAAGGTCATGACTATCCTAGAACTCTCTGTGGTTTTTGAAAATC 1440

Qy      1506 ATTGAATTC 1514
Db       1441 ATTGAATTC 1449
```

RESULT 5
US-09-333-729A-8
; Sequence 8, Application US/09333729A
; Patent No. 6270987
; GENERAL INFORMATION:
; APPLICANT: Wang,Yang
; APPLICANT: Spellman,Michael W.
; TITLE OF INVENTION: O-Fucosyltransferase
; FILE REFERENCE: P1041P1D1-Substitute
; CURRENT APPLICATION NUMBER: US/09/333,729A
; CURRENT FILING DATE: 1999-06-15
; PRIOR APPLICATION NUMBER: US 08/798,741
; PRIOR FILING DATE: 1997-11-26
; NUMBER OF SEQ ID NOS: 21
; SEQ ID NO 8
; LENGTH: 5009
; TYPE: DNA
; ORGANISM: Homo Sapien
US-09-333-729A-8

Query Match 95.6%; Score 1447.4; DB 3; Length 5009;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 1448; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```
Qy         66 GAACCAGGCCGATCACTTCTTGGGCTCTCTGGCATTTGCAAAGCTGCTAAACCGTACCTT 125
Db          1 GAACCAGGCCGATCACTTCTTGGGCTCTCTGGCATTTGCAAAGCTGCTAAACCGTACCTT 60

Qy        126 GGCTGTCCCTCCTTGGATTGAGTACCAGCATCACAAGCCTCCTTTCACCAACCTCCATGT 185
Db         61 GGCTGTCCCTCCTTGGATTGAGTACCAGCATCACAAGCCTCCTTTCACCAACCTCCATGT 120

Qy        186 GTCCTACCAGAAGTACTTCAAGCTGGAGCCCCTCCAGGCTTACCATCGGGTCATCAGCTT 245
Db        121 GTCCTACCAGAAGTACTTCAAGCTGGAGCCCCTCCAGGCTTACCATCGGGTCATCAGCTT 180

Qy        246 GGAGGATTTCATGGAGAAGCTGGCACCCACCCACTGGCCCCCTGAGAAGCGGGTGGCATA 305
Db        181 GGAGGATTTCATGGAGAAGCTGGCACCCACCCACTGGCCCCCTGAGAAGCGGGTGGCATA 240

Qy        306 CTGCTTTGAGGTGGCAGCCCAGCGAAGCCCAGATAAGAAGACGTGCCCCATGAAGGAAGG 365
Db        241 CTGCTTTGAGGTGGCAGCCCAGCGAAGCCCAGATAAGAAGACGTGCCCCATGAAGGAAGG 300

Qy        366 AAACCCCTTTGGCCCATTCTGGGATCAGTTTCATGTGAGTTTCAACAAGTCGGAGCTTTT 425
Db        301 AAACCCCTTTGGCCCATTCTGGGATCAGTTTCATGTGAGTTTCAACAAGTCGGAGCTTTT 360

Qy        426 TACAGGCATTTCCTTCAGTGCTTCCTACAGAGAACAATGGAGCCAGAGATTTTCTCCAAA 485
Db        361 TACAGGCATTTCCTTCAGTGCTTCCTACAGAGAACAATGGAGCCAGAGATTTTCTCCAAA 420

Qy        486 GGAACATCCGGTGCTTGCCCTGCCAGGAGCCCCAGCCCAGTTCCCCGTCCTAGAAGAACA 545
Db        421 GGAACATCCGGTGCTTGCCCTGCCAGGAGCCCCAGCCCAGTTCCCCGTCCTAGAGGAACA 480
```

```
Qy       546 CAGGCCACTACAGAAGTACATGGTATGGTCAGACGAAATGGTGAAGACGGGAGAGGCCCA 605
Db       481 CAGGCCACTACAGAAGTACATGGTATGGTCAGACGAAATGGTGAAGACGGGAGAGGCCCA 540

Qy       606 GATTCATGCCCACCTTGTCCGGCCCTATGTGGGCATTCATCTGCGCATTGGCTCTGACTG 665
Db       541 GATTCATGCCCACCTTGTCCGGCCCTATGTGGGCATTCATCTGCGCATTGGCTCTGACTG 600

Qy       666 GAAGAACGCCTGTGCCATGCTGAAGGACGGGACTGCAGGCTCGCACTTCATGGCCTCTCC 725
Db       601 GAAGAACGCCTGTGCCATGCTGAAGGACGGGACTGCAGGCTCGCACTTCATGGCCTCTCC 660

Qy       726 GCAGTGTGTGGGCTACAGCCGCAGCACAGCGGCCCCCCTCACGATGACTATGTGCCTGCC 785
Db       661 GCAGTGTGTGGGCTACAGCCGCAGCACAGCGGCCCCCCTCACGATGACTATGTGCCTGCC 720

Qy       786 TGACCTGAAGGAGATCCAGAGGGCTGTGAAGCTCTGGGTGAGGTCGCTGGATGCCCAGTC 845
Db       721 TGACCTGAAGGAGATCCAGAGGGCTGTGAAGCTCTGGGTGAGGTCGCTGGATGCCCAGTC 780

Qy       846 GGTCTACGTTGCTACTGATTCCGAGAGTTATGTGCCTGAGCTCCAACAGCTCTTCAAAGG 905
Db       781 GGTCTACGTTGCTACTGATTCCGAGAGTTATGTGCCTGAGCTCCAACAGCTCTTCAAAGG 840

Qy       906 GAAGGTGAAGGTGGTGAGCCTGAAGCCTGAGGTGGCCCAGGTCGACCTGTACATCCTCGG 965
Db       841 GAAGGTGAAGGTGGTGAGCCTGAAGCCTGAGGTGGCCCAGGTCGACCTGTACATCCTCGG 900

Qy       966 CCAAGCCGACCACTTTATTGGCAACTGTGTCTCCTCCTTCACTGCCTTTGTGAAGCGGGA 1025
Db       901 CCAAGCCGACCACTTTATTGGCAACTGTGTCTCCTCCTTCACTGCCTTTGTGAAGCGGGA 960

Qy      1026 GCGGGACCTCCAGGGGAGGCCGTCTTCTTTCTTCGGCATGGACAGGCCCCCTAAGCTGCG 1085
Db       961 GCGGGACCTCCAGGGGAGGCCGTCTTCTTTCTTCGGCATGGACAGGCCCCCTAAGCTGCG 1020

Qy      1086 GGACGAGTTCTGATTCTGGCCGGAGCACCAGACCCTCTGATCCTGGAGGGACCAGAGTCT 1145
Db      1021 GGACGAGTTCTGATTCTGGCCGGAGCACCAGACCCTCTGATCCTGGAGGGACCAGAGTCT 1080

Qy      1146 GAGCTGGTCCTTCCAGCCAGGCCTGGCAGCCAGAGGTGCTCCGGGATTGCAAACTCCTCT 1205
Db      1081 GAGCTGGTCCTTCCAGCCAGGCCTGGCAGCCAGAGGTGCTCCGGGATTGCAAACTCCTCT 1140

Qy      1206 TCTCACCTGCCAAAGATGGAGAAGAGTGCCAGGGACCCCTCAAGGAGGGAGACGCTCCAT 1265
Db      1141 TCTCACCTGCCAAAGATGGAGAAGAGTGCCAGGGACCCCTCAAGGAGGGAGACGCTCCAT 1200

Qy      1266 ATCCCAGGGCATAGGACTTGCAGGTTCCTAGGAGCAGGAGCATCTCCCATCGCACGTGCT 1325
Db      1201 ATCCCAGGGCATAGGACTTGCAGGTTCCTAGGAGCAGGAGCATCTCCCATCGCACGTGCT 1260

Qy      1326 TTCTGCTCTTCTGGGAATTTCTCACACTGGCAAAGCAGTCCAGCCTCCGTCTTCTGGTCC 1385
Db      1261 TTCTGCTCTTCTGGGAATTTCTCACACTGGCAAAGCAGTCCAGCCTCCGTCTTCTGGTCC 1320

Qy      1386 ACTCTGCTCTGAGCAGCCTGGGATGCTGAACTCTTCAGAGAGATTTTTTTATAGAGAGAT 1445
Db      1321 ACTCTGCTCTGAGCAGCCTGGGATGCTGAACTCTTCAGAGAGATTTTTTTATAGAGAGAT 1380

Qy      1446 TTCTATAATTTTGATACAAGGTCATGACTATCCTAGAACTCTCTGTGGTTTTTGAAAATC 1505
Db      1381 TTCTATAATTTTGATACAAGGTCATGACTATCCTAGAACTCTCTGTGGTTTTTGAAAATC 1440

Qy      1506 ATTGAATTC 1514
Db      1441 ATTGAATTC 1449


RESULT 6
US-08-978-741-4
; Sequence 4, Application US/08978741
; Patent No. 6100076
;  GENERAL INFORMATION:
;    APPLICANT:  Yang Wang, Michael W. Spellman
;    TITLE OF INVENTION:  O-Fucosyltransferase
;    NUMBER OF SEQUENCES:  17
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Genentech, Inc.
;      STREET:  1 DNA Way
;      CITY:  South San Francisco
;      STATE:  California
;      COUNTRY:  USA
;      ZIP:  94080
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  3.5 inch, 1.44 Mb floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  WinPatin (Genentech)
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/978,741
;      FILING DATE:  26-No. 6100076-1997
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/792498
;      FILING DATE:  31
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Svoboda, Craig G.
;      REGISTRATION NUMBER:  39,044
;      REFERENCE/DOCKET NUMBER:  P1041P1
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  650/225-1489
;      TELEFAX:  650/952-9881
;  INFORMATION FOR SEQ ID NO:  4:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1300 base pairs
;      TYPE:  Nucleic Acid
;      STRANDEDNESS:  Single
;      TOPOLOGY:  Linear
US-08-978-741-4

  Query Match             76.8%;  Score 1163.4;  DB 3;  Length 1300;
  Best Local Similarity   99.9%;  Pred. No. 0;
  Matches 1164;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 ATGCCCGCGGGCTCCTGGGACCCGGCCGGTTACCTGCTCTACTGCCCCTGCATGGGGCGC 60
Db        136 ATGCCCGCGGGCTCCTGGGACCCGGCCGGTTACCTGCTCTACTGCCCCTGCATGGGGCGC 195

Qy         61 TTTGGGAACCAGGCCGATCACTTCTTGGGCTCTCTGGCATTTGCAAAGCTGCTAAACCGT 120
Db        196 TTTGGGAACCAGGCCGATCACTTCTTGGGCTCTCTGGCATTTGCAAAGCTGCTAAACCGT 255

Qy        121 ACCTTGGCTGTCCCTCCTTGGATTGAGTACCAGCATCACAAGCCTCCTTTCACCAACCTC 180
Db        256 ACCTTGGCTGTCCCTCCTTGGATTGAGTACCAGCATCACAAGCCTCCTTTCACCAACCTC 315

Qy        181 CATGTGTCCTACCAGAAGTACTTCAAGCTGGAGCCCCTCCAGGCTTACCATCGGGTCATC 240
Db        316 CATGTGTCCTACCAGAAGTACTTCAAGCTGGAGCCCCTCCAGGCTTACCATCGGGTCATC 375

Qy        241 AGCTTGGAGGATTTCATGGAGAAGCTGGCACCCACCCACTGGCCCCCTGAGAAGCGGGTG 300
Db        376 AGCTTGGAGGATTTCATGGAGAAGCTGGCACCCACCCACTGGCCCCCTGAGAAGCGGGTG 435

Qy        301 GCATACTGCTTTGAGGTGGCAGCCCAGCGAAGCCCAGATAAGAAGACGTGCCCCATGAAG 360
Db        436 GCATACTGCTTTGAGGTGGCAGCCCAGCGAAGCCCAGATAAGAAGACGTGCCCCATGAAG 495

Qy        361 GAAGGAAACCCCTTTGGCCCATTCTGGGATCAGTTTCATGTGAGTTTCAACAAGTCGGAG 420
Db        496 GAAGGAAACCCCTTTGGCCCATTCTGGGATCAGTTTCATGTGAGTTTCAACAAGTCGGAG 555

Qy        421 CTTTTTACAGGCATTTCCTTCAGTGCTTCCTACAGAGAACAATGGAGCCAGAGATTTTCT 480
Db        556 CTTTTTACAGGCATTTCCTTCAGTGCTTCCTACAGAGAACAATGGAGCCAGAGATTTTCT 615
```

```
Qy        481 CCAAAGGAACATCCGGTGCTTGCCCTGCCAGGAGCCCCAGCCCAGTTCCCCGTCCTAGAA 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        616 CCAAAGGAACATCCGGTGCTTGCCCTGCCAGGAGCCCCAGCCCAGTTCCCCGTCCTAGAG 675

Qy        541 GAACACAGGCCACTACAGAAGTACATGGTATGGTCAGACGAAATGGTGAAGACGGGAGAG 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        676 GAACACAGGCCACTACAGAAGTACATGGTATGGTCAGACGAAATGGTGAAGACGGGAGAG 735

Qy        601 GCCCAGATTCATGCCCACCTTGTCCGGCCCTATGTGGGCATTCATCTGCGCATTGGCTCT 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        736 GCCCAGATTCATGCCCACCTTGTCCGGCCCTATGTGGGCATTCATCTGCGCATTGGCTCT 795

Qy        661 GACTGGAAGAACGCCTGTGCCATGCTGAAGGACGGGACTGCAGGCTCGCACTTCATGGCC 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        796 GACTGGAAGAACGCCTGTGCCATGCTGAAGGACGGGACTGCAGGCTCGCACTTCATGGCC 855

Qy        721 TCTCCGCAGTGTGTGGGCTACAGCCGCAGCACAGCGGCCCCCCTCACGATGACTATGTGC 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        856 TCTCCGCAGTGTGTGGGCTACAGCCGCAGCACAGCGGCCCCCCTCACGATGACTATGTGC 915

Qy        781 CTGCCTGACCTGAAGGAGATCCAGAGGGCTGTGAAGCTCTGGGTGAGGTCGCTGGATGCC 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        916 CTGCCTGACCTGAAGGAGATCCAGAGGGCTGTGAAGCTCTGGGTGAGGTCGCTGGATGCC 975

Qy        841 CAGTCGGTCTACGTTGCTACTGATTCCGAGAGTTATGTGCCTGAGCTCCAACAGCTCTTC 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        976 CAGTCGGTCTACGTTGCTACTGATTCCGAGAGTTATGTGCCTGAGCTCCAACAGCTCTTC 1035

Qy        901 AAAGGGAAGGTGAAGGTGGTGAGCCTGAAGCCTGAGGTGGCCCAGGTCGACCTGTACATC 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1036 AAAGGGAAGGTGAAGGTGGTGAGCCTGAAGCCTGAGGTGGCCCAGGTCGACCTGTACATC 1095

Qy        961 CTCGGCCAAGCCGACCACTTTATTGGCAACTGTGTCTCCTCCTTCACTGCCTTTGTGAAG 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1096 CTCGGCCAAGCCGACCACTTTATTGGCAACTGTGTCTCCTCCTTCACTGCCTTTGTGAAG 1155

Qy       1021 CGGGAGCGGGACCTCCAGGGGAGGCCGTCTTCTTTCTTCGGCATGGACAGGCCCCCTAAG 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1156 CGGGAGCGGGACCTCCAGGGGAGGCCGTCTTCTTTCTTCGGCATGGACAGGCCCCCTAAG 1215

Qy       1081 CTGCGGGACGAGTTCTGATTCTGGCCGGAGCACCAGACCCTCTGATCCTGGAGGGACCAG 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1216 CTGCGGGACGAGTTCTGATTCTGGCCGGAGCACCAGACCCTCTGATCCTGGAGGGACCAG 1275

Qy       1141 AGTCTGAGCTGGTCCTTCCAGCCAG 1165
              |||||||||||||||||||||||||
Db       1276 AGTCTGAGCTGGTCCTTCCAGCCAG 1300
```

RESULT 7
US-09-333-729A-6
; Sequence 6, Application US/09333729A
; Patent No. 6270987
; GENERAL INFORMATION:
; APPLICANT: Wang,Yang
; APPLICANT: Spellman,Michael W.
; TITLE OF INVENTION: O-Fucosyltransferase
; FILE REFERENCE: P1041P1D1-Substitute
; CURRENT APPLICATION NUMBER: US/09/333,729A
; CURRENT FILING DATE: 1999-06-15
; PRIOR APPLICATION NUMBER: US 08/798,741
; PRIOR FILING DATE: 1997-11-26
; NUMBER OF SEQ ID NOS: 21
; SEQ ID NO 6
; LENGTH: 1300
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Plasmid insert.
US-09-333-729A-6

Query Match 76.8%; Score 1163.4; DB 3; Length 1300;

Best Local Similarity 99.9%; Pred. No. 0;
Matches 1164; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```
Qy          1 ATGCCCGCGGGCTCCTGGGACCCGGCCGGTTACCTGCTCTACTGCCCCTGCATGGGGCGC 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        136 ATGCCCGCGGGCTCCTGGGACCCGGCCGGTTACCTGCTCTACTGCCCCTGCATGGGGCGC 195

Qy         61 TTTGGGAACCAGGCCGATCACTTCTTGGGCTCTCTGGCATTTGCAAAGCTGCTAAACCGT 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        196 TTTGGGAACCAGGCCGATCACTTCTTGGGCTCTCTGGCATTTGCAAAGCTGCTAAACCGT 255

Qy        121 ACCTTGGCTGTCCCTCCTTGGATTGAGTACCAGCATCACAAGCCTCCTTTCACCAACCTC 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        256 ACCTTGGCTGTCCCTCCTTGGATTGAGTACCAGCATCACAAGCCTCCTTTCACCAACCTC 315

Qy        181 CATGTGTCCTACCAGAAGTACTTCAAGCTGGAGCCCCTCCAGGCTTACCATCGGGTCATC 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        316 CATGTGTCCTACCAGAAGTACTTCAAGCTGGAGCCCCTCCAGGCTTACCATCGGGTCATC 375

Qy        241 AGCTTGGAGGATTTCATGGAGAAGCTGGCACCCACCCACTGGCCCCCTGAGAAGCGGGTG 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        376 AGCTTGGAGGATTTCATGGAGAAGCTGGCACCCACCCACTGGCCCCCTGAGAAGCGGGTG 435

Qy        301 GCATACTGCTTTGAGGTGGCAGCCCAGCGAAGCCCAGATAAGAAGACGTGCCCCATGAAG 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        436 GCATACTGCTTTGAGGTGGCAGCCCAGCGAAGCCCAGATAAGAAGACGTGCCCCATGAAG 495

Qy        361 GAAGGAAACCCCTTTGGCCCATTCTGGGATCAGTTTCATGTGAGTTTCAACAAGTCGGAG 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        496 GAAGGAAACCCCTTTGGCCCATTCTGGGATCAGTTTCATGTGAGTTTCAACAAGTCGGAG 555

Qy        421 CTTTTTACAGGCATTTCCTTCAGTGCTTCCTACAGAGAACAATGGAGCCAGAGATTTTCT 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        556 CTTTTTACAGGCATTTCCTTCAGTGCTTCCTACAGAGAACAATGGAGCCAGAGATTTTCT 615

Qy        481 CCAAAGGAACATCCGGTGCTTGCCCTGCCAGGAGCCCCAGCCCAGTTCCCCGTCCTAGAA 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        616 CCAAAGGAACATCCGGTGCTTGCCCTGCCAGGAGCCCCAGCCCAGTTCCCCGTCCTAGAG 675

Qy        541 GAACACAGGCCACTACAGAAGTACATGGTATGGTCAGACGAAATGGTGAAGACGGGAGAG 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        676 GAACACAGGCCACTACAGAAGTACATGGTATGGTCAGACGAAATGGTGAAGACGGGAGAG 735

Qy        601 GCCCAGATTCATGCCCACCTTGTCCGGCCCTATGTGGGCATTCATCTGCGCATTGGCTCT 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        736 GCCCAGATTCATGCCCACCTTGTCCGGCCCTATGTGGGCATTCATCTGCGCATTGGCTCT 795

Qy        661 GACTGGAAGAACGCCTGTGCCATGCTGAAGGACGGGACTGCAGGCTCGCACTTCATGGCC 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        796 GACTGGAAGAACGCCTGTGCCATGCTGAAGGACGGGACTGCAGGCTCGCACTTCATGGCC 855

Qy        721 TCTCCGCAGTGTGTGGGCTACAGCCGCAGCACAGCGGCCCCCCTCACGATGACTATGTGC 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        856 TCTCCGCAGTGTGTGGGCTACAGCCGCAGCACAGCGGCCCCCCTCACGATGACTATGTGC 915

Qy        781 CTGCCTGACCTGAAGGAGATCCAGAGGGCTGTGAAGCTCTGGGTGAGGTCGCTGGATGCC 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        916 CTGCCTGACCTGAAGGAGATCCAGAGGGCTGTGAAGCTCTGGGTGAGGTCGCTGGATGCC 975

Qy        841 CAGTCGGTCTACGTTGCTACTGATTCCGAGAGTTATGTGCCTGAGCTCCAACAGCTCTTC 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        976 CAGTCGGTCTACGTTGCTACTGATTCCGAGAGTTATGTGCCTGAGCTCCAACAGCTCTTC 1035

Qy        901 AAAGGGAAGGTGAAGGTGGTGAGCCTGAAGCCTGAGGTGGCCCAGGTCGACCTGTACATC 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1036 AAAGGGAAGGTGAAGGTGGTGAGCCTGAAGCCTGAGGTGGCCCAGGTCGACCTGTACATC 1095

Qy        961 CTCGGCCAAGCCGACCACTTTATTGGCAACTGTGTCTCCTCCTTCACTGCCTTTGTGAAG 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1096 CTCGGCCAAGCCGACCACTTTATTGGCAACTGTGTCTCCTCCTTCACTGCCTTTGTGAAG 1155

Qy       1021 CGGGAGCGGGACCTCCAGGGGAGGCCGTCTTCTTTCTTCGGCATGGACAGGCCCCCTAAG 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db       1156 CGGGAGCGGGACCTCCAGGGGAGGCCGTCTTCTTTCTTCGGCATGGACAGGCCCCCTAAG 1215

Qy       1081 CTGCGGGACGAGTTCTGATTCTGGCCGGAGCACCAGACCCTCTGATCCTGGAGGGACCAG 1140
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1216 CTGCGGGACGAGTTCTGATTCTGGCCGGAGCACCAGACCCTCTGATCCTGGAGGGACCAG 1275

Qy       1141 AGTCTGAGCTGGTCCTTCCAGCCAG 1165
              |||||||||||||||||||||||||
Db       1276 AGTCTGAGCTGGTCCTTCCAGCCAG 1300


RESULT 8
US-08-978-741-16
; Sequence 16, Application US/08978741
; Patent No. 6100076
;  GENERAL INFORMATION:
;    APPLICANT:  Yang Wang, Michael W. Spellman
;    TITLE OF INVENTION:  O-Fucosyltransferase
;    NUMBER OF SEQUENCES:  17
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Genentech, Inc.
;      STREET:  1 DNA Way
;      CITY:  South San Francisco
;      STATE:  California
;      COUNTRY:  USA
;      ZIP:  94080
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  3.5 inch, 1.44 Mb floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  WinPatin (Genentech)
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/978,741
;      FILING DATE:  26-No. 6100076-1997
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/792498
;      FILING DATE:  31
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Svoboda, Craig G.
;      REGISTRATION NUMBER:  39,044
;      REFERENCE/DOCKET NUMBER:  P1041P1
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  650/225-1489
;      TELEFAX:  650/952-9881
;  INFORMATION FOR SEQ ID NO:  16:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1100 base pairs
;      TYPE:  Nucleic Acid
;      STRANDEDNESS:  Single
;      TOPOLOGY:  Linear
US-08-978-741-16

  Query Match             72.7%;  Score 1100;  DB 3;  Length 1100;
  Best Local Similarity   100.0%;  Pred. No. 0;
  Matches 1100;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 ATGCCCGCGGGCTCCTGGGACCCGGCCGGTTACCTGCTCTACTGCCCCTGCATGGGGCGC 60
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGCCCGCGGGCTCCTGGGACCCGGCCGGTTACCTGCTCTACTGCCCCTGCATGGGGCGC 60

Qy         61 TTTGGGAACCAGGCCGATCACTTCTTGGGCTCTCTGGCATTTGCAAAGCTGCTAAACCGT 120
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 TTTGGGAACCAGGCCGATCACTTCTTGGGCTCTCTGGCATTTGCAAAGCTGCTAAACCGT 120

Qy        121 ACCTTGGCTGTCCCTCCTTGGATTGAGTACCAGCATCACAAGCCTCCTTTCACCAACCTC 180
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 ACCTTGGCTGTCCCTCCTTGGATTGAGTACCAGCATCACAAGCCTCCTTTCACCAACCTC 180

Qy        181 CATGTGTCCTACCAGAAGTACTTCAAGCTGGAGCCCCTCCAGGCTTACCATCGGGTCATC 240
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 CATGTGTCCTACCAGAAGTACTTCAAGCTGGAGCCCCTCCAGGCTTACCATCGGGTCATC 240

Qy        241 AGCTTGGAGGATTTCATGGAGAAGCTGGCACCCACCCACTGGCCCCCTGAGAAGCGGGTG 300
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 AGCTTGGAGGATTTCATGGAGAAGCTGGCACCCACCCACTGGCCCCCTGAGAAGCGGGTG 300

Qy        301 GCATACTGCTTTGAGGTGGCAGCCCAGCGAAGCCCAGATAAGAAGACGTGCCCCATGAAG 360
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 GCATACTGCTTTGAGGTGGCAGCCCAGCGAAGCCCAGATAAGAAGACGTGCCCCATGAAG 360

Qy        361 GAAGGAAACCCCTTTGGCCCATTCTGGGATCAGTTTCATGTGAGTTTCAACAAGTCGGAG 420
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 GAAGGAAACCCCTTTGGCCCATTCTGGGATCAGTTTCATGTGAGTTTCAACAAGTCGGAG 420

Qy        421 CTTTTTACAGGCATTTCCTTCAGTGCTTCCTACAGAGAACAATGGAGCCAGAGATTTTCT 480
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 CTTTTTACAGGCATTTCCTTCAGTGCTTCCTACAGAGAACAATGGAGCCAGAGATTTTCT 480

Qy        481 CCAAAGGAACATCCGGTGCTTGCCCTGCCAGGAGCCCCAGCCCAGTTCCCCGTCCTAGAA 540
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 CCAAAGGAACATCCGGTGCTTGCCCTGCCAGGAGCCCCAGCCCAGTTCCCCGTCCTAGAA 540

Qy        541 GAACACAGGCCACTACAGAAGTACATGGTATGGTCAGACGAAATGGTGAAGACGGGAGAG 600
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 GAACACAGGCCACTACAGAAGTACATGGTATGGTCAGACGAAATGGTGAAGACGGGAGAG 600

Qy        601 GCCCAGATTCATGCCCACCTTGTCCGGCCCTATGTGGGCATTCATCTGCGCATTGGCTCT 660
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 GCCCAGATTCATGCCCACCTTGTCCGGCCCTATGTGGGCATTCATCTGCGCATTGGCTCT 660

Qy        661 GACTGGAAGAACGCCTGTGCCATGCTGAAGGACGGGACTGCAGGCTCGCACTTCATGGCC 720
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 GACTGGAAGAACGCCTGTGCCATGCTGAAGGACGGGACTGCAGGCTCGCACTTCATGGCC 720

Qy        721 TCTCCGCAGTGTGTGGGCTACAGCCGCAGCACAGCGGCCCCCCTCACGATGACTATGTGC 780
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 TCTCCGCAGTGTGTGGGCTACAGCCGCAGCACAGCGGCCCCCCTCACGATGACTATGTGC 780

Qy        781 CTGCCTGACCTGAAGGAGATCCAGAGGGCTGTGAAGCTCTGGGTGAGGTCGCTGGATGCC 840
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 CTGCCTGACCTGAAGGAGATCCAGAGGGCTGTGAAGCTCTGGGTGAGGTCGCTGGATGCC 840

Qy        841 CAGTCGGTCTACGTTGCTACTGATTCCGAGAGTTATGTGCCTGAGCTCCAACAGCTCTTC 900
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 CAGTCGGTCTACGTTGCTACTGATTCCGAGAGTTATGTGCCTGAGCTCCAACAGCTCTTC 900

Qy        901 AAAGGGAAGGTGAAGGTGGTGAGCCTGAAGCCTGAGGTGGCCCAGGTCGACCTGTACATC 960
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 AAAGGGAAGGTGAAGGTGGTGAGCCTGAAGCCTGAGGTGGCCCAGGTCGACCTGTACATC 960

Qy        961 CTCGGCCAAGCCGACCACTTTATTGGCAACTGTGTCTCCTCCTTCACTGCCTTTGTGAAG 1020
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 CTCGGCCAAGCCGACCACTTTATTGGCAACTGTGTCTCCTCCTTCACTGCCTTTGTGAAG 1020

Qy       1021 CGGGAGCGGGACCTCCAGGGGAGGCCGTCTTCTTTCTTCGGCATGGACAGGCCCCCTAAG 1080
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1021 CGGGAGCGGGACCTCCAGGGGAGGCCGTCTTCTTTCTTCGGCATGGACAGGCCCCCTAAG 1080

Qy       1081 CTGCGGGACGAGTTCTGATT 1100
              ||||||||||||||||||||
Db       1081 CTGCGGGACGAGTTCTGATT 1100


RESULT 9
US-09-333-729A-16
; Sequence 16, Application US/09333729A
; Patent No. 6270987
; GENERAL INFORMATION:
;  APPLICANT:  Wang,Yang
;  APPLICANT:  Spellman,Michael W.
;  TITLE OF INVENTION:  O-Fucosyltransferase
;  FILE REFERENCE:  P1041P1D1-Substitute
;  CURRENT APPLICATION NUMBER: US/09/333,729A
;  CURRENT FILING DATE:  1999-06-15
;  PRIOR APPLICATION NUMBER:  US 08/798,741
```

```
;  PRIOR FILING DATE:  1997-11-26
;  NUMBER OF SEQ ID NOS:  21
; SEQ ID NO 16
;   LENGTH: 1100
;   TYPE: DNA
;   ORGANISM: Homo Sapien
US-09-333-729A-16

  Query Match             72.7%;  Score 1100;  DB 3;  Length 1100;
  Best Local Similarity   100.0%;  Pred. No. 0;
  Matches 1100;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 ATGCCCGCGGGCTCCTGGGACCCGGCCGGTTACCTGCTCTACTGCCCCTGCATGGGGCGC 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGCCCGCGGGCTCCTGGGACCCGGCCGGTTACCTGCTCTACTGCCCCTGCATGGGGCGC 60

Qy         61 TTTGGGAACCAGGCCGATCACTTCTTGGGCTCTCTGGCATTTGCAAAGCTGCTAAACCGT 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 TTTGGGAACCAGGCCGATCACTTCTTGGGCTCTCTGGCATTTGCAAAGCTGCTAAACCGT 120

Qy        121 ACCTTGGCTGTCCCTCCTTGGATTGAGTACCAGCATCACAAGCCTCCTTTCACCAACCTC 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 ACCTTGGCTGTCCCTCCTTGGATTGAGTACCAGCATCACAAGCCTCCTTTCACCAACCTC 180

Qy        181 CATGTGTCCTACCAGAAGTACTTCAAGCTGGAGCCCCTCCAGGCTTACCATCGGGTCATC 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 CATGTGTCCTACCAGAAGTACTTCAAGCTGGAGCCCCTCCAGGCTTACCATCGGGTCATC 240

Qy        241 AGCTTGGAGGATTTCATGGAGAAGCTGGCACCCACCCACTGGCCCCCTGAGAAGCGGGTG 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 AGCTTGGAGGATTTCATGGAGAAGCTGGCACCCACCCACTGGCCCCCTGAGAAGCGGGTG 300

Qy        301 GCATACTGCTTTGAGGTGGCAGCCCAGCGAAGCCCAGATAAGAAGACGTGCCCCATGAAG 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 GCATACTGCTTTGAGGTGGCAGCCCAGCGAAGCCCAGATAAGAAGACGTGCCCCATGAAG 360

Qy        361 GAAGGAAACCCCTTTGGCCCATTCTGGGATCAGTTTCATGTGAGTTTCAACAAGTCGGAG 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 GAAGGAAACCCCTTTGGCCCATTCTGGGATCAGTTTCATGTGAGTTTCAACAAGTCGGAG 420

Qy        421 CTTTTTACAGGCATTTCCTTCAGTGCTTCCTACAGAGAACAATGGAGCCAGAGATTTTCT 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 CTTTTTACAGGCATTTCCTTCAGTGCTTCCTACAGAGAACAATGGAGCCAGAGATTTTCT 480

Qy        481 CCAAAGGAACATCCGGTGCTTGCCCTGCCAGGAGCCCCAGCCCAGTTCCCCGTCCTAGAA 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 CCAAAGGAACATCCGGTGCTTGCCCTGCCAGGAGCCCCAGCCCAGTTCCCCGTCCTAGAA 540

Qy        541 GAACACAGGCCACTACAGAAGTACATGGTATGGTCAGACGAAATGGTGAAGACGGGAGAG 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 GAACACAGGCCACTACAGAAGTACATGGTATGGTCAGACGAAATGGTGAAGACGGGAGAG 600

Qy        601 GCCCAGATTCATGCCCACCTTGTCCGGCCCTATGTGGGCATTCATCTGCGCATTGGCTCT 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 GCCCAGATTCATGCCCACCTTGTCCGGCCCTATGTGGGCATTCATCTGCGCATTGGCTCT 660

Qy        661 GACTGGAAGAACGCCTGTGCCATGCTGAAGGACGGGACTGCAGGCTCGCACTTCATGGCC 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 GACTGGAAGAACGCCTGTGCCATGCTGAAGGACGGGACTGCAGGCTCGCACTTCATGGCC 720

Qy        721 TCTCCGCAGTGTGTGGGCTACAGCCGCAGCACAGCGGCCCCCCTCACGATGACTATGTGC 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 TCTCCGCAGTGTGTGGGCTACAGCCGCAGCACAGCGGCCCCCCTCACGATGACTATGTGC 780

Qy        781 CTGCCTGACCTGAAGGAGATCCAGAGGGCTGTGAAGCTCTGGGTGAGGTCGCTGGATGCC 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 CTGCCTGACCTGAAGGAGATCCAGAGGGCTGTGAAGCTCTGGGTGAGGTCGCTGGATGCC 840

Qy        841 CAGTCGGTCTACGTTGCTACTGATTCCGAGAGTTATGTGCCTGAGCTCCAACAGCTCTTC 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 CAGTCGGTCTACGTTGCTACTGATTCCGAGAGTTATGTGCCTGAGCTCCAACAGCTCTTC 900

Qy        901 AAAGGGAAGGTGAAGGTGGTGAGCCTGAAGCCTGAGGTGGCCCAGGTCGACCTGTACATC 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 AAAGGGAAGGTGAAGGTGGTGAGCCTGAAGCCTGAGGTGGCCCAGGTCGACCTGTACATC 960

Qy        961 CTCGGCCAAGCCGACCACTTTATTGGCAACTGTGTCTCCTCCTTCACTGCCTTTGTGAAG 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 CTCGGCCAAGCCGACCACTTTATTGGCAACTGTGTCTCCTCCTTCACTGCCTTTGTGAAG 1020

Qy       1021 CGGGAGCGGGACCTCCAGGGGAGGCCGTCTTCTTTCTTCGGCATGGACAGGCCCCCTAAG 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1021 CGGGAGCGGGACCTCCAGGGGAGGCCGTCTTCTTTCTTCGGCATGGACAGGCCCCCTAAG 1080

Qy       1081 CTGCGGGACGAGTTCTGATT 1100
              ||||||||||||||||||||
Db       1081 CTGCGGGACGAGTTCTGATT 1100


RESULT 10
US-09-270-767-14353
; Sequence 14353, Application US/09270767
; Patent No. 6703491
; GENERAL INFORMATION:
;  APPLICANT: Homburger et al.
;  TITLE OF INVENTION: Nucleic acids and proteins of Drosophila melanogaster
;  FILE REFERENCE: File Reference: 7326-094
;  CURRENT APPLICATION NUMBER: US/09/270,767
;  CURRENT FILING DATE: 1999-03-17
;  NUMBER OF SEQ ID NOS: 62517
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 14353
;   LENGTH: 1320
;   TYPE: DNA
;   ORGANISM: Drosophila melanogaster
US-09-270-767-14353

  Query Match             11.1%;  Score 168;  DB 4;  Length 1320;
  Best Local Similarity   51.6%;  Pred. No. 6.1e-41;
  Matches  564;  Conservative    0;  Mismatches  485;  Indels   45;  Gaps    6;

Qy         17 GGGACCCGGCCGGTTACCTGCTCTACTGCCCCTGCATGGGGCGCTTTGGGAACCAGGCCG 76
              | || ||    || |||||  ||||||| || || ||||| |||||||| ||||||||||
Db         73 GCGATCCCAATGGCTACCTCACCTACTGTCCGTGTATGGGACGCTTTGGCAACCAGGCCG 132

Qy         77 ATCACTTCTTGGGCTCTCTGGCATTTGCAAAGCTGCTAAACCGTACCTTGGCTGTCCCTC 136
              | |||||| |||| ||  |||| || || ||| ||| || || ||| ||    | || |
Db        133 ACCACTTCCTGGGATCATTGGCCTTCGCCAAGGCGCTTAATCGCACCCTGATCCTGCCGC 192

Qy        137 CTTGGATTGAGTACCAGCATCACAAGCCTCCTTTCACCAACCTCCATGTGTCCTACCAGA 196
              | ||| | ||||       ||| | |  |    |   || || ||  | | | | |
Db        193 CGTGGGTGGAGT------ATCGTAGGGGTGAACTGCGATCCCGTCAGGTACCGTTCAACA 246

Qy        197 AGTACTTCAAGCTGGAGCCCCTCCAGGCTTACCATCGGGTCATCAGCTTGGAGGATTTCA 256
              || ||  |  |||||||||  |||  |||||||| |||||| | |||  |||||||
Db        247 CATATTTTGAAGTGGAGCCCCTGAAGGAATACCATCGCGTCATCACCATGGCAGATTTCA 306

Qy        257 TGGAGAAGCTGGCACCCACCCACTGGCCCCCTGAGAAGCGGGTGGCATACTGCTTTGAGG 316
              ||  | | |||||   |  |   |||||       | |||| ||| |||  ||||   |||
Db        307 TGTGGCACCTGGCCGACGACATTTGGCCAGAATCGGAGCGAGTGTCATTTTGCTACAAGG 366

Qy        317 ---------------TGGCAGCCCAGCGAAGCCCAGATAAGAAGACGTGCCCCATGAAGG 361
                            |  |   ||     ||||| |||   |  |||| |   ||||
Db        367 AACGATATAGCCTTCAGCAGGAGAAGAACGATCCAGACAAGCCCAATTGCCACGCCAAGG 426

Qy        362 AAGGAAACCCCTTTGGCCCATTCTGGGATCAGTTTCATGTGAGTTTCAACAAGTCGGAGC 421
              | || || || ||||| || || |||||    |||||  |    ||       ||| |||
Db        427 ATGGCAATCCTTTTGGTCCCTTTTGGGACACTTTTCACATAGACTTTGTGCGGTCAGAGT 486

Qy        422 TTTTTACAGGCATTTCCTTCAGTGCTTCCTACAG---------AGAACAATGGAGCCAGA 472
              | | | |    ||   ||  ||    | ||          |   |  | |||      |
Db        487 TCTATGCGCCACTTCATTTTGATGTGCATCATAGCAACGAGGCTGCCAAGTGGCAGACCA 546

Qy        473 GATTTTCTCCAAAGGAACATCCGGTGCTTGCCCTGCCAGGAGCCCCAGCCCAGTTCCCCG 532
              || | || || |   | |||| || || ||  |  | ||||| || ||    || || |
Db        547 AATATCCTGCAGAATCATATCCCGTACTCGCGTTCACCGGAGCTCCGGCTAGTTTTCCTG 606
```

```
Qy        533 TCCTAGAAGAACACAGGCCACTACAGAAGTACATGGTATGGTCAGACGAAATGGTGAAGA 592
Db        607 TTCAGCTAGAGAACTGCAAGCTGCAGCGCTACTTGCAGTGGAGTCAACGGTATAGGGAAG 666

Qy        593 CGGGAGAGGCCCAGATTCATGCCCACCTTGTCCGGC---CCTATGTGGGCATTCATCTGC 649
Db        667 CATCTAAGGATTTCATCCGAGAGCAGTTGCCTCGGGGTGCCTTTTTGGGCATTCATCTGC 726

Qy        650 GCATTGGCTCTGACTGGAAGAACGCCTGTGCCATGCTGAAGGACGG---------GACTG 700
Db        727 GCAACGGTATCGATTGGGTGAGAGCCTGTGAGCACGTCAAGGATAGCCAGCATCTGTTTG 786

Qy        701 CAGGCTCGCACTTCATGGCCTCTCCGCAGTGTGTGGGCTACAGCCGCAGCACAGCGGCCC 760
Db        787 CCTCGCCGCAGTGCTTGGGCTATAAAAATGAACGTGGTGCACTCTACCCGGAGCTCTGCA 846

Qy        761 CCCTCACGATGACTATGTGCCTGCCTGACCTGAAGGAGATCCAGAGGGCTGTGAAGCTCT 820
Db        847 TGCCCTCCAAGGAGGCGATCATCCGCCAGCTAAAGAGAACCATTAAGAACGTGCGCCAAA 906

Qy        821 GGGTGAGGTCGCTGGATGCCCAGTCGGTCTACGTTGCTACTGATTCCGAGAGTTATGTGC 880
Db        907 CTCAGCCGGACAACGAAATCAAATCAGTTTTCGTGGCGTCAGACTCCAATCACATGATTG 966

Qy        881 CTGAGCTCCAACAGCTCTTCAAAGGGAAGG---TGAAGGTGGTGAGCCTGAAGCCTGAGG 937
Db        967 GTGAACTAAACACGGCCCTTAGTCGCATGGGCATCAGTGTGCACAAGCTGACGGAGGATG 1026

Qy        938 TGGCCCAGGTCGACCTGTACATCCTCGGCCAAGCCGACCACTTTATTGGCAACTGTGTCT 997
Db       1027 ATCCTTACCTGGACTTGGCCATTCTCGGACAGTCGAACCACTTTATCGGCAACTGTATAT 1086

Qy        998 CCTCCTTCACTGCCTTTGTGAAGCGGGAGCGGGACCTCCAGGGGAGGCCGTCTTCTTTCT 1057
Db       1087 CCTCTTACTCGGCATTCGTAAAAAGGGAACGAGATGTGCACGGTTTTCCATCGTACTTCT 1146

Qy       1058 TCGGCATGGACAGG 1071
Db       1147 GGGGATTCCCCAAG 1160
```

RESULT 11
US-09-270-767-30507
; Sequence 30507, Application US/09270767
; Patent No. 6703491
; GENERAL INFORMATION:
; APPLICANT: Homburger et al.
; TITLE OF INVENTION: Nucleic acids and proteins of Drosophila melanogaster
; FILE REFERENCE: File Reference: 7326-094
; CURRENT APPLICATION NUMBER: US/09/270,767
; CURRENT FILING DATE: 1999-03-17
; NUMBER OF SEQ ID NOS: 62517
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 30507
; LENGTH: 722
; TYPE: DNA
; ORGANISM: Drosophila melanogaster
US-09-270-767-30507

Query Match 4.1%; Score 62.2; DB 4; Length 722;
Best Local Similarity 49.5%; Pred. No. 1.7e-08;
Matches 225; Conservative 0; Mismatches 218; Indels 12; Gaps 2;

```
Qy        629 CCTATGTGGGCATTCATCTGCGCATTGGCTCTGACTGGAAGAACGCCTGTGCCATGCTGA 688
Db        108 CCTTTTTGGGCATTCATCTGCGCAACGGTATCGATTGGGTGAGAGCCTGTGAGCACGTCA 167

Qy        689 AGGACGG---------GACTGCAGGCTCGCACTTCATGGCCTCTCCGCAGTGTGTGGGCT 739
Db        168 AGGATAGCCAGCATCTGTTTGCCTCGCCGCAGTGCTTGGGCTATAAAAATGAACGTGGTG 227

Qy        740 ACAGCCGCAGCACAGCGGCCCCCCTCACGATGACTATGTGCCTGCCTGACCTGAAGGAGA 799
Db        228 CACTCTACCCGGAGCTCTGCATGCCCTCCAAGGAGGCGATCATCCGCCAGCTAAAGAGAA 287

Qy        800 TCCAGAGGGCTGTGAAGCTCTGGGTGAGGTCGCTGGATGCCCAGTCGGTCTACGTTGCTA 859
Db        288 CCATTAAGAACGTGCGCCAAACTCAGCCGGACAACGAAATCAAATCAGTTTTCGTGGCGT 347

Qy        860 CTGATTCCGAGAGTTATGTGCCTGAGCTCCAACAGCTCTTCAAAGGGAAGG---TGAAGG 916
Db        348 CAGACTCCAATCACATGATTGGTGAACTAAACACGGCCCTTAGTCGCATGGGCATCAGTG 407

Qy        917 TGGTGAGCCTGAAGCCTGAGGTGGCCCAGGTCGACCTGTACATCCTCGGCCAAGCCGACC 976
Db        408 TGCACAAGCTGACGGAGGATGATCCTTACCTGGACTTGGCCATTCTCGGACAGTCGAACC 467

Qy        977 ACTTTATTGGCAACTGTGTCTCCTCCTTCACTGCCTTTGTGAAGCGGGAGCGGGACCTCC 1036
Db        468 ACTTTATCGGCAACTGTATATCCTCTTACTCGGCATTCGTAAAAAGGGAACGAGATGTGC 527

Qy       1037 AGGGGAGGCCGTCTTCTTTCTTCGGCATGGACAGG 1071
Db        528 ACGGTTTTCCATCGTACTTCTGGGGATTCCCCAAG 562
```

RESULT 12
US-09-270-767-28905/c
; Sequence 28905, Application US/09270767
; Patent No. 6703491
; GENERAL INFORMATION:
; APPLICANT: Homburger et al.
; TITLE OF INVENTION: Nucleic acids and proteins of Drosophila melanogaster
; FILE REFERENCE: File Reference: 7326-094
; CURRENT APPLICATION NUMBER: US/09/270,767
; CURRENT FILING DATE: 1999-03-17
; NUMBER OF SEQ ID NOS: 62517
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 28905
; LENGTH: 998
; TYPE: DNA
; ORGANISM: Drosophila melanogaster
US-09-270-767-28905

Query Match 3.9%; Score 58.6; DB 4; Length 998;
Best Local Similarity 60.2%; Pred. No. 2.7e-07;
Matches 97; Conservative 0; Mismatches 64; Indels 0; Gaps 0;

```
Qy        911 TGAAGGTGGTGAGCCTGAAGCCTGAGGTGGCCCAGGTCGACCTGTACATCCTCGGCCAAG 970
Db        896 TCAGTGTGCACAAGCTGACGGAGGATGATCCTTACCTGGACTTGGCCATTCTCGGACAGT 837

Qy        971 CCGACCACTTTATTGGCAACTGTGTCTCCTCCTTCACTGCCTTTGTGAAGCGGGAGCGGG 1030
Db        836 CGAACCACTTTATCGGCAACTGTATATCCTCTTACTCGGCATTCGTAAAAAGGGAACGAG 777

Qy       1031 ACCTCCAGGGGAGGCCGTCTTCTTTCTTCGGCATGGACAGG 1071
Db        776 ATGTGCACGGTTTTCCATCGTACTTCTGGGGATTCCCCAAG 736
```

RESULT 13
US-09-270-767-13022/c
; Sequence 13022, Application US/09270767
; Patent No. 6703491
; GENERAL INFORMATION:
; APPLICANT: Homburger et al.
; TITLE OF INVENTION: Nucleic acids and proteins of Drosophila melanogaster
; FILE REFERENCE: File Reference: 7326-094
; CURRENT APPLICATION NUMBER: US/09/270,767
; CURRENT FILING DATE: 1999-03-17
; NUMBER OF SEQ ID NOS: 62517
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 13022
; LENGTH: 2034

```
;   TYPE: DNA
;   ORGANISM: Drosophila melanogaster
US-09-270-767-13022

  Query Match             3.9%;  Score 58.6;  DB 4;  Length 2034;
  Best Local Similarity   60.2%;  Pred. No. 4.4e-07;
  Matches   97;  Conservative    0;  Mismatches   64;  Indels    0;  Gaps    0;

Qy        911 TGAAGGTGGTGAGCCTGAAGCCTGAGGTGGCCCAGGTCGACCTGTACATCCTCGGCCAAG 970
              | |  |||    |  |||| |   || |    |  |  | ||| ||  ||| ||||| ||
Db       1932 TCAGTGTGCACAAGCTGACGGAGGATGATCCTTACCTGGACTTGGCCATTCTCGGACAGT 1873

Qy        971 CCGACCACTTTATTGGCAACTGTGTCTCCTCCTTCACTGCCTTTGTGAAGCGGGAGCGGG 1030
              |  |||||||||| |||||||| | ||||| | | | || || || ||  |||| || |
Db       1872 CGAACCACTTTATCGGCAACTGTATATCCTCTTACTCGGCATTCGTAAAAAGGGAACGAG 1813

Qy       1031 ACCTCCAGGGGAGGCCGTCTTCTTTCTTCGGCATGGACAGG 1071
              |  | || ||    || || |  ||||  ||  |   || |
Db       1812 ATGTGCACGGTTTTCCATCGTACTTCTGGGGATTCCCCAAG 1772


RESULT 14
US-09-949-016-12127/c
; Sequence 12127, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
;  APPLICANT: VENTER, J. Craig et al.
;  TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
;  TITLE OF INVENTION:  WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
;  FILE REFERENCE: CL001307
;  CURRENT APPLICATION NUMBER: US/09/949,016
;  CURRENT FILING DATE:  2000-04-14
;  PRIOR APPLICATION NUMBER: 60/241,755
;  PRIOR FILING DATE: 2000-10-20
;  PRIOR APPLICATION NUMBER: 60/237,768
;  PRIOR FILING DATE: 2000-10-03
;  PRIOR APPLICATION NUMBER: 60/231,498
;  PRIOR FILING DATE: 2000-09-08
;  NUMBER OF SEQ ID NOS: 207012
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 12127
;   LENGTH: 19227
;   TYPE: DNA
;   ORGANISM: Human
US-09-949-016-12127

  Query Match             3.7%;  Score 56;  DB 4;  Length 19227;
  Best Local Similarity   100.0%;  Pred. No. 1.2e-05;
  Matches   56;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 ATGCCCGCGGGCTCCTGGGACCCGGCCGGTTACCTGCTCTACTGCCCCTGCATGGG 56
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1718 ATGCCCGCGGGCTCCTGGGACCCGGCCGGTTACCTGCTCTACTGCCCCTGCATGGG 1663


RESULT 15
US-09-949-016-16285/c
; Sequence 16285, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
;  APPLICANT: VENTER, J. Craig et al.
;  TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
;  TITLE OF INVENTION:  WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
;  FILE REFERENCE: CL001307
;  CURRENT APPLICATION NUMBER: US/09/949,016
;  CURRENT FILING DATE:  2000-04-14
;  PRIOR APPLICATION NUMBER: 60/241,755
;  PRIOR FILING DATE: 2000-10-20
;  PRIOR APPLICATION NUMBER: 60/237,768
;  PRIOR FILING DATE: 2000-10-03
;  PRIOR APPLICATION NUMBER: 60/231,498
;  PRIOR FILING DATE: 2000-09-08
;  NUMBER OF SEQ ID NOS: 207012
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 16285
;   LENGTH: 19228
;   TYPE: DNA
;   ORGANISM: Human
US-09-949-016-16285

  Query Match             3.7%;  Score 56;  DB 4;  Length 19228;
  Best Local Similarity   100.0%;  Pred. No. 1.2e-05;
  Matches   56;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 ATGCCCGCGGGCTCCTGGGACCCGGCCGGTTACCTGCTCTACTGCCCCTGCATGGG 56
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1718 ATGCCCGCGGGCTCCTGGGACCCGGCCGGTTACCTGCTCTACTGCCCCTGCATGGG 1663


Search completed: November  9, 2005, 10:46:05
Job time : 279.956 secs
```

GenCore version 5.1.6

OM nucleic - nucleic search, using sw model

Run on: November 9, 2005, 08:00:04 ; Search time 1355.3 Seconds
(without alignments)
9238.165 Million cell updates/sec

Title: US-09-774-954-1
Perfect score: 1514
Sequence: 1 ATGCCCGCGGGCTCCTGGGA.....TTTTGAAAATCATTGAATTC 1514

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 9794790 seqs, 4134909567 residues

Total number of hits satisfying chosen parameters: 19589580

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published_Applications_NA:*
1: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq:*
2: /cgn2_6/ptodata/1/pubpna/PCT_NEW_PUB.seq:*
3: /cgn2_6/ptodata/1/pubpna/US06_NEW_PUB.seq:*
4: /cgn2_6/ptodata/1/pubpna/US06_PUBCOMB.seq:*
5: /cgn2_6/ptodata/1/pubpna/US07_NEW_PUB.seq:*
6: /cgn2_6/ptodata/1/pubpna/PCTUS_PUBCOMB.seq:*
7: /cgn2_6/ptodata/1/pubpna/US08_NEW_PUB.seq:*
8: /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq:*
9: /cgn2_6/ptodata/1/pubpna/US09A_PUBCOMB.seq:*
10: /cgn2_6/ptodata/1/pubpna/US09B_PUBCOMB.seq:*
11: /cgn2_6/ptodata/1/pubpna/US09C_PUBCOMB.seq:*
12: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq:*
13: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq2:*
14: /cgn2_6/ptodata/1/pubpna/US10A_PUBCOMB.seq:*
15: /cgn2_6/ptodata/1/pubpna/US10B_PUBCOMB.seq:*
16: /cgn2_6/ptodata/1/pubpna/US10C_PUBCOMB.seq:*
17: /cgn2_6/ptodata/1/pubpna/US10D_PUBCOMB.seq:*
18: /cgn2_6/ptodata/1/pubpna/US10E_PUBCOMB.seq:*
19: /cgn2_6/ptodata/1/pubpna/US10F_PUBCOMB.seq:*
20: /cgn2_6/ptodata/1/pubpna/US10G_PUBCOMB.seq:*
21: /cgn2_6/ptodata/1/pubpna/US10H_PUBCOMB.seq:*
22: /cgn2_6/ptodata/1/pubpna/US10I_PUBCOMB.seq:*
23: /cgn2_6/ptodata/1/pubpna/US10_NEW_PUB.seq:*
24: /cgn2_6/ptodata/1/pubpna/US10_NEW_PUB.seq2:*
25: /cgn2_6/ptodata/1/pubpna/US11A_PUBCOMB.seq:*
26: /cgn2_6/ptodata/1/pubpna/US11_NEW_PUB.seq:*
27: /cgn2_6/ptodata/1/pubpna/US60_NEW_PUB.seq:*
28: /cgn2_6/ptodata/1/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	1514	100.0	1514	11	US-09-774-954-1	Sequence 1, Appli
	2	1512.4	99.9	11284	11	US-09-774-954-5	Sequence 5, Appli
	3	1510.8	99.8	5218	16	US-10-301-822-168	Sequence 168, App
	4	1510.8	99.8	5218	22	US-10-956-157-4752	Sequence 4752, Ap
	5	1447.4	95.6	5009	11	US-09-774-954-7	Sequence 7, Appli
	6	1163.4	76.8	1300	11	US-09-774-954-4	Sequence 4, Appli
	7	1100	72.7	1100	11	US-09-774-954-16	Sequence 16, Appl
	8	608.6	40.2	10331	10	US-09-764-891-6032	Sequence 6032, Ap
	9	471.4	31.1	1400	22	US-10-956-157-9986	Sequence 9986, Ap
	10	471.4	31.1	1506	22	US-10-956-157-4751	Sequence 4751, Ap
	11	442.8	29.2	477	10	US-09-918-995-22744	Sequence 22744, A
	12	441	29.1	479	10	US-09-918-995-19437	Sequence 19437, A
	13	166.4	11.0	1209	26	US-11-097-143-18161	Sequence 18161, A
c	14	150.4	9.9	3264	26	US-11-097-143-18160	Sequence 18160, A
	15	114.6	7.6	402	9	US-09-783-590-11501	Sequence 11501, A
c	16	108.6	7.2	3793	26	US-11-097-143-18082	Sequence 18082, A
c	17	71.8	4.7	3567	26	US-11-097-143-18169	Sequence 18169, A
	18	60	4.0	60	10	US-09-908-975-17287	Sequence 17287, A
	19	56	3.7	760	24	US-10-820-474A-236	Sequence 236, App
c	20	55.6	3.7	521	13	US-09-925-065A-602444	Sequence 602444,
	21	42.4	2.8	1438	18	US-10-282-122A-19431	Sequence 19431, A
c	22	41.2	2.7	114793	16	US-10-148-806-3	Sequence 3, Appli
c	23	41.2	2.7	114793	24	US-10-859-792-3	Sequence 3, Appli
	24	40.6	2.7	2036	18	US-10-369-493-43154	Sequence 43154, A
	25	40.4	2.7	1212	21	US-10-425-115-114097	Sequence 114097,
	26	40.4	2.7	5932	16	US-10-205-823-312	Sequence 312, App
	27	40.4	2.7	5932	26	US-11-051-454-312	Sequence 312, App
	28	40.2	2.7	726	20	US-10-437-963-19394	Sequence 19394, A
	29	40	2.6	40	11	US-09-774-954-10	Sequence 10, Appl
	30	40	2.6	40	11	US-09-774-954-12	Sequence 12, Appl
	31	40	2.6	2032	20	US-10-437-963-8098	Sequence 8098, Ap
c	32	39.2	2.6	618	13	US-09-925-065A-571171	Sequence 571171,
	33	39	2.6	372	18	US-10-282-122A-31476	Sequence 31476, A
	34	39	2.6	1209	10	US-09-948-820-37	Sequence 37, Appl
	35	39	2.6	1209	21	US-10-613-076-37	Sequence 37, Appl
	36	39	2.6	1555	24	US-10-220-335-120	Sequence 120, App
	37	39	2.6	1575	19	US-10-072-012-131	Sequence 131, App
	38	39	2.6	1654	24	US-10-220-335-121	Sequence 121, App
c	39	39	2.6	1963	24	US-10-220-335-464	Sequence 464, App
c	40	39	2.6	1963	24	US-10-220-335-465	Sequence 465, App
	41	38.6	2.5	407	9	US-09-960-352-11732	Sequence 11732, A
	42	38.6	2.5	2358	10	US-09-919-039-190	Sequence 190, App
	43	38.6	2.5	2384	22	US-10-278-698-123	Sequence 123, App
	44	38.6	2.5	2384	22	US-10-278-698-638	Sequence 638, App
	45	38.6	2.5	2670	17	US-10-247-671-64	Sequence 64, Appl

ALIGNMENTS

RESULT 1
US-09-774-954-1
; Sequence 1, Application US/09774954
; Publication No. US20040241645A1
; GENERAL INFORMATION:
; APPLICANT: Yang Wang, Michael W. Spellman
; TITLE OF INVENTION: O-Fucosyltransferase
; NUMBER OF SEQUENCES: 17
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genentech, Inc.
; STREET: 1 DNA Way
; CITY: South San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94080
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5 inch, 1.44 Mb floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: WinPatin (Genentech)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/774,954
; FILING DATE: 30-Jan-2001
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/978,741
; FILING DATE: 26-NOV-1997
; APPLICATION NUMBER: 08/792,498

```
;            FILING DATE: 31-JAN-1997
;       ATTORNEY/AGENT INFORMATION:
;            NAME: Svoboda, Craig G.
;            REGISTRATION NUMBER: 39,044
;            REFERENCE/DOCKET NUMBER: P1041P1
;       TELECOMMUNICATION INFORMATION:
;            TELEPHONE: 650/225-1489
;            TELEFAX: 650/952-9881
;   INFORMATION FOR SEQ ID NO: 1:
;       SEQUENCE CHARACTERISTICS:
;            LENGTH: 1514 base pairs
;            TYPE: Nucleic Acid
;            STRANDEDNESS: Single
;            TOPOLOGY: Linear
;       SEQUENCE DESCRIPTION: SEQ ID NO: 1:
US-09-774-954-1

  Query Match            100.0%;  Score 1514;  DB 11;  Length 1514;
  Best Local Similarity  100.0%;  Pred. No. 0;
  Matches 1514;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 ATGCCCGCGGGCTCCTGGGACCCGGCCGGTTACCTGCTCTACTGCCCCTGCATGGGGCGC 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGCCCGCGGGCTCCTGGGACCCGGCCGGTTACCTGCTCTACTGCCCCTGCATGGGGCGC 60

Qy         61 TTTGGGAACCAGGCCGATCACTTCTTGGGCTCTCTGGCATTTGCAAAGCTGCTAAACCGT 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 TTTGGGAACCAGGCCGATCACTTCTTGGGCTCTCTGGCATTTGCAAAGCTGCTAAACCGT 120

Qy        121 ACCTTGGCTGTCCCTCCTTGGATTGAGTACCAGCATCACAAGCCTCCTTTCACCAACCTC 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 ACCTTGGCTGTCCCTCCTTGGATTGAGTACCAGCATCACAAGCCTCCTTTCACCAACCTC 180

Qy        181 CATGTGTCCTACCAGAAGTACTTCAAGCTGGAGCCCCTCCAGGCTTACCATCGGGTCATC 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 CATGTGTCCTACCAGAAGTACTTCAAGCTGGAGCCCCTCCAGGCTTACCATCGGGTCATC 240

Qy        241 AGCTTGGAGGATTTCATGGAGAAGCTGGCACCCACCCACTGGCCCCCTGAGAAGCGGGTG 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 AGCTTGGAGGATTTCATGGAGAAGCTGGCACCCACCCACTGGCCCCCTGAGAAGCGGGTG 300

Qy        301 GCATACTGCTTTGAGGTGGCAGCCCAGCGAAGCCCAGATAAGAAGACGTGCCCCATGAAG 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 GCATACTGCTTTGAGGTGGCAGCCCAGCGAAGCCCAGATAAGAAGACGTGCCCCATGAAG 360

Qy        361 GAAGGAAACCCCTTTGGCCCATTCTGGGATCAGTTTCATGTGAGTTTCAACAAGTCGGAG 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 GAAGGAAACCCCTTTGGCCCATTCTGGGATCAGTTTCATGTGAGTTTCAACAAGTCGGAG 420

Qy        421 CTTTTTACAGGCATTTCCTTCAGTGCTTCCTACAGAGAACAATGGAGCCAGAGATTTTCT 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 CTTTTTACAGGCATTTCCTTCAGTGCTTCCTACAGAGAACAATGGAGCCAGAGATTTTCT 480

Qy        481 CCAAAGGAACATCCGGTGCTTGCCCTGCCAGGAGCCCCAGCCCAGTTCCCCGTCCTAGAA 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 CCAAAGGAACATCCGGTGCTTGCCCTGCCAGGAGCCCCAGCCCAGTTCCCCGTCCTAGAA 540

Qy        541 GAACACAGGCCACTACAGAAGTACATGGTATGGTCAGACGAAATGGTGAAGACGGGAGAG 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 GAACACAGGCCACTACAGAAGTACATGGTATGGTCAGACGAAATGGTGAAGACGGGAGAG 600

Qy        601 GCCCAGATTCATGCCCACCTTGTCCGGCCCTATGTGGGCATTCATCTGCGCATTGGCTCT 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 GCCCAGATTCATGCCCACCTTGTCCGGCCCTATGTGGGCATTCATCTGCGCATTGGCTCT 660

Qy        661 GACTGGAAGAACGCCTGTGCCATGCTGAAGGACGGGACTGCAGGCTCGCACTTCATGGCC 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 GACTGGAAGAACGCCTGTGCCATGCTGAAGGACGGGACTGCAGGCTCGCACTTCATGGCC 720

Qy        721 TCTCCGCAGTGTGTGGGCTACAGCCGCAGCACAGCGGCCCCCCTCACGATGACTATGTGC 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 TCTCCGCAGTGTGTGGGCTACAGCCGCAGCACAGCGGCCCCCCTCACGATGACTATGTGC 780

Qy        781 CTGCCTGACCTGAAGGAGATCCAGAGGGCTGTGAAGCTCTGGGTGAGGTCGCTGGATGCC 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 CTGCCTGACCTGAAGGAGATCCAGAGGGCTGTGAAGCTCTGGGTGAGGTCGCTGGATGCC 840

Qy        841 CAGTCGGTCTACGTTGCTACTGATTCCGAGAGTTATGTGCCTGAGCTCCAACAGCTCTTC 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 CAGTCGGTCTACGTTGCTACTGATTCCGAGAGTTATGTGCCTGAGCTCCAACAGCTCTTC 900

Qy        901 AAAGGGAAGGTGAAGGTGGTGAGCCTGAAGCCTGAGGTGGCCCAGGTCGACCTGTACATC 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 AAAGGGAAGGTGAAGGTGGTGAGCCTGAAGCCTGAGGTGGCCCAGGTCGACCTGTACATC 960

Qy        961 CTCGGCCAAGCCGACCACTTTATTGGCAACTGTGTCTCCTCCTTCACTGCCTTTGTGAAG 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 CTCGGCCAAGCCGACCACTTTATTGGCAACTGTGTCTCCTCCTTCACTGCCTTTGTGAAG 1020

Qy       1021 CGGGAGCGGGACCTCCAGGGGAGGCCGTCTTCTTTCTTCGGCATGGACAGGCCCCCTAAG 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1021 CGGGAGCGGGACCTCCAGGGGAGGCCGTCTTCTTTCTTCGGCATGGACAGGCCCCCTAAG 1080

Qy       1081 CTGCGGGACGAGTTCTGATTCTGGCCGGAGCACCAGACCCTCTGATCCTGGAGGGACCAG 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1081 CTGCGGGACGAGTTCTGATTCTGGCCGGAGCACCAGACCCTCTGATCCTGGAGGGACCAG 1140

Qy       1141 AGTCTGAGCTGGTCCTTCCAGCCAGGCCTGGCAGCCAGAGGTGCTCCGGGATTGCAAACT 1200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1141 AGTCTGAGCTGGTCCTTCCAGCCAGGCCTGGCAGCCAGAGGTGCTCCGGGATTGCAAACT 1200

Qy       1201 CCTCTTCTCACCTGCCAAAGATGGAGAAGAGTGCCAGGGACCCCTCAAGGAGGGAGACGC 1260
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1201 CCTCTTCTCACCTGCCAAAGATGGAGAAGAGTGCCAGGGACCCCTCAAGGAGGGAGACGC 1260

Qy       1261 TCCATATCCCAGGGCATAGGACTTGCAGGTTCCTAGGAGCAGGAGCATCTCCCATCGCAC 1320
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1261 TCCATATCCCAGGGCATAGGACTTGCAGGTTCCTAGGAGCAGGAGCATCTCCCATCGCAC 1320

Qy       1321 GTGCTTTCTGCTCTTCTGGGAATTTCTCACACTGGCAAAGCAGTCCAGCCTCCGTCTTCT 1380
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1321 GTGCTTTCTGCTCTTCTGGGAATTTCTCACACTGGCAAAGCAGTCCAGCCTCCGTCTTCT 1380

Qy       1381 GGTCCACTCTGCTCTGAGCAGCCTGGGATGCTGAACTCTTCAGAGAGATTTTTTTATAGA 1440
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1381 GGTCCACTCTGCTCTGAGCAGCCTGGGATGCTGAACTCTTCAGAGAGATTTTTTTATAGA 1440

Qy       1441 GAGATTTCTATAATTTTGATACAAGGTCATGACTATCCTAGAACTCTCTGTGGTTTTTGA 1500
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1441 GAGATTTCTATAATTTTGATACAAGGTCATGACTATCCTAGAACTCTCTGTGGTTTTTGA 1500

Qy       1501 AAATCATTGAATTC 1514
              ||||||||||||||
Db       1501 AAATCATTGAATTC 1514


RESULT 2
US-09-774-954-5
; Sequence 5, Application US/09774954
; Publication No. US20040241645A1
;   GENERAL INFORMATION:
;        APPLICANT: Yang Wang, Michael W. Spellman
;        TITLE OF INVENTION: O-Fucosyltransferase
;        NUMBER OF SEQUENCES: 17
;        CORRESPONDENCE ADDRESS:
;             ADDRESSEE: Genentech, Inc.
;             STREET: 1 DNA Way
;             CITY: South San Francisco
;             STATE: California
;             COUNTRY: USA
;             ZIP: 94080
;        COMPUTER READABLE FORM:
;             MEDIUM TYPE: 3.5 inch, 1.44 Mb floppy disk
;             COMPUTER: IBM PC compatible
;             OPERATING SYSTEM: PC-DOS/MS-DOS
;             SOFTWARE: WinPatin (Genentech)
```

```
;       CURRENT APPLICATION DATA:
;            APPLICATION NUMBER: US/09/774,954
;            FILING DATE: 30-Jan-2001
;            CLASSIFICATION: <Unknown>
;       PRIOR APPLICATION DATA:
;            APPLICATION NUMBER: US/08/978,741
;            FILING DATE: 26-NOV-1997
;            APPLICATION NUMBER: 08/792,498
;            FILING DATE: 31-JAN-1997
;       ATTORNEY/AGENT INFORMATION:
;            NAME: Svoboda, Craig G.
;            REGISTRATION NUMBER: 39,044
;            REFERENCE/DOCKET NUMBER: P1041P1
;       TELECOMMUNICATION INFORMATION:
;            TELEPHONE: 650/225-1489
;            TELEFAX: 650/952-9881
;   INFORMATION FOR SEQ ID NO: 5:
;       SEQUENCE CHARACTERISTICS:
;            LENGTH: 11284 base pairs
;            TYPE: Nucleic Acid
;            STRANDEDNESS: Single
;            TOPOLOGY: Linear
;       SEQUENCE DESCRIPTION: SEQ ID NO: 5:
US-09-774-954-5

  Query Match           99.9%;  Score 1512.4;  DB 11;  Length 11284;
  Best Local Similarity 99.9%;  Pred. No. 0;
  Matches 1513;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 ATGCCCGCGGGCTCCTGGGACCCGGCCGGTTACCTGCTCTACTGCCCCTGCATGGGGCGC 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4236 ATGCCCGCGGGCTCCTGGGACCCGGCCGGTTACCTGCTCTACTGCCCCTGCATGGGGCGC 4295

Qy         61 TTTGGGAACCAGGCCGATCACTTCTTGGGCTCTCTGGCATTTGCAAAGCTGCTAAACCGT 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4296 TTTGGGAACCAGGCCGATCACTTCTTGGGCTCTCTGGCATTTGCAAAGCTGCTAAACCGT 4355

Qy        121 ACCTTGGCTGTCCCTCCTTGGATTGAGTACCAGCATCACAAGCCTCCTTTCACCAACCTC 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4356 ACCTTGGCTGTCCCTCCTTGGATTGAGTACCAGCATCACAAGCCTCCTTTCACCAACCTC 4415

Qy        181 CATGTGTCCTACCAGAAGTACTTCAAGCTGGAGCCCCTCCAGGCTTACCATCGGGTCATC 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4416 CATGTGTCCTACCAGAAGTACTTCAAGCTGGAGCCCCTCCAGGCTTACCATCGGGTCATC 4475

Qy        241 AGCTTGGAGGATTTCATGGAGAAGCTGGCACCCACCCACTGGCCCCCTGAGAAGCGGGTG 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4476 AGCTTGGAGGATTTCATGGAGAAGCTGGCACCCACCCACTGGCCCCCTGAGAAGCGGGTG 4535

Qy        301 GCATACTGCTTTGAGGTGGCAGCCCAGCGAAGCCCAGATAAGAAGACGTGCCCCATGAAG 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4536 GCATACTGCTTTGAGGTGGCAGCCCAGCGAAGCCCAGATAAGAAGACGTGCCCCATGAAG 4595

Qy        361 GAAGGAAACCCCTTTGGCCCATTCTGGGATCAGTTTCATGTGAGTTTCAACAAGTCGGAG 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4596 GAAGGAAACCCCTTTGGCCCATTCTGGGATCAGTTTCATGTGAGTTTCAACAAGTCGGAG 4655

Qy        421 CTTTTTACAGGCATTTCCTTCAGTGCTTCCTACAGAGAACAATGGAGCCAGAGATTTTCT 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4656 CTTTTTACAGGCATTTCCTTCAGTGCTTCCTACAGAGAACAATGGAGCCAGAGATTTTCT 4715

Qy        481 CCAAAGGAACATCCGGTGCTTGCCCTGCCAGGAGCCCCAGCCCAGTTCCCCGTCCTAGAA 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4716 CCAAAGGAACATCCGGTGCTTGCCCTGCCAGGAGCCCCAGCCCAGTTCCCCGTCCTAGAG 4775

Qy        541 GAACACAGGCCACTACAGAAGTACATGGTATGGTCAGACGAAATGGTGAAGACGGGAGAG 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4776 GAACACAGGCCACTACAGAAGTACATGGTATGGTCAGACGAAATGGTGAAGACGGGAGAG 4835

Qy        601 GCCCAGATTCATGCCCACCTTGTCCGGCCCTATGTGGGCATTCATCTGCGCATTGGCTCT 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4836 GCCCAGATTCATGCCCACCTTGTCCGGCCCTATGTGGGCATTCATCTGCGCATTGGCTCT 4895

Qy        661 GACTGGAAGAACGCCTGTGCCATGCTGAAGGACGGGACTGCAGGCTCGCACTTCATGGCC 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4896 GACTGGAAGAACGCCTGTGCCATGCTGAAGGACGGGACTGCAGGCTCGCACTTCATGGCC 4955

Qy        721 TCTCCGCAGTGTGTGGGCTACAGCCGCAGCACAGCGGCCCCCCTCACGATGACTATGTGC 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4956 TCTCCGCAGTGTGTGGGCTACAGCCGCAGCACAGCGGCCCCCCTCACGATGACTATGTGC 5015

Qy        781 CTGCCTGACCTGAAGGAGATCCAGAGGGCTGTGAAGCTCTGGGTGAGGTCGCTGGATGCC 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5016 CTGCCTGACCTGAAGGAGATCCAGAGGGCTGTGAAGCTCTGGGTGAGGTCGCTGGATGCC 5075

Qy        841 CAGTCGGTCTACGTTGCTACTGATTCCGAGAGTTATGTGCCTGAGCTCCAACAGCTCTTC 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5076 CAGTCGGTCTACGTTGCTACTGATTCCGAGAGTTATGTGCCTGAGCTCCAACAGCTCTTC 5135

Qy        901 AAAGGGAAGGTGAAGGTGGTGAGCCTGAAGCCTGAGGTGGCCCAGGTCGACCTGTACATC 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5136 AAAGGGAAGGTGAAGGTGGTGAGCCTGAAGCCTGAGGTGGCCCAGGTCGACCTGTACATC 5195

Qy        961 CTCGGCCAAGCCGACCACTTTATTGGCAACTGTGTCTCCTCCTTCACTGCCTTTGTGAAG 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5196 CTCGGCCAAGCCGACCACTTTATTGGCAACTGTGTCTCCTCCTTCACTGCCTTTGTGAAG 5255

Qy       1021 CGGGAGCGGGACCTCCAGGGGAGGCCGTCTTCTTTCTTCGGCATGGACAGGCCCCCTAAG 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5256 CGGGAGCGGGACCTCCAGGGGAGGCCGTCTTCTTTCTTCGGCATGGACAGGCCCCCTAAG 5315

Qy       1081 CTGCGGGACGAGTTCTGATTCTGGCCGGAGCACCAGACCCTCTGATCCTGGAGGGACCAG 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5316 CTGCGGGACGAGTTCTGATTCTGGCCGGAGCACCAGACCCTCTGATCCTGGAGGGACCAG 5375

Qy       1141 AGTCTGAGCTGGTCCTTCCAGCCAGGCCTGGCAGCCAGAGGTGCTCCGGGATTGCAAACT 1200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5376 AGTCTGAGCTGGTCCTTCCAGCCAGGCCTGGCAGCCAGAGGTGCTCCGGGATTGCAAACT 5435

Qy       1201 CCTCTTCTCACCTGCCAAAGATGGAGAAGAGTGCCAGGGACCCCTCAAGGAGGGAGACGC 1260
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5436 CCTCTTCTCACCTGCCAAAGATGGAGAAGAGTGCCAGGGACCCCTCAAGGAGGGAGACGC 5495

Qy       1261 TCCATATCCCAGGGCATAGGACTTGCAGGTTCCTAGGAGCAGGAGCATCTCCCATCGCAC 1320
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5496 TCCATATCCCAGGGCATAGGACTTGCAGGTTCCTAGGAGCAGGAGCATCTCCCATCGCAC 5555

Qy       1321 GTGCTTTCTGCTCTTCTGGGAATTTCTCACACTGGCAAAGCAGTCCAGCCTCCGTCTTCT 1380
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5556 GTGCTTTCTGCTCTTCTGGGAATTTCTCACACTGGCAAAGCAGTCCAGCCTCCGTCTTCT 5615

Qy       1381 GGTCCACTCTGCTCTGAGCAGCCTGGGATGCTGAACTCTTCAGAGAGATTTTTTTATAGA 1440
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5616 GGTCCACTCTGCTCTGAGCAGCCTGGGATGCTGAACTCTTCAGAGAGATTTTTTTATAGA 5675

Qy       1441 GAGATTTCTATAATTTTGATACAAGGTCATGACTATCCTAGAACTCTCTGTGGTTTTTGA 1500
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5676 GAGATTTCTATAATTTTGATACAAGGTCATGACTATCCTAGAACTCTCTGTGGTTTTTGA 5735

Qy       1501 AAATCATTGAATTC 1514
              ||||||||||||||
Db       5736 AAATCATTGAATTC 5749


RESULT 3
US-10-301-822-168
; Sequence 168, Application US/10301822
; Publication No. US20030148410A1
; GENERAL INFORMATION:
;  APPLICANT: Millennium Pharmaceuticals, Inc.
;  APPLICANT:  Berger, Allison
;  APPLICANT:  Guillemette, Tracy L.
;  APPLICANT:  Kamatkar, Shubhangi
;  APPLICANT:  Schlegel, Robert
;  APPLICANT:  Monahan, John E.
;  APPLICANT:  Thibodeau, Stephen N.
```

```
;  APPLICANT:  Burgart, Lawrence J.
;  TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND
;  TITLE OF INVENTION:  METHODS FOR IDENTIFICATION, ASSESSMENT, PREVENTION, AND
;  TITLE OF INVENTION:  THERAPY OF COLON CANCER
;  FILE REFERENCE: MPM01-029P2RNM
;  CURRENT APPLICATION NUMBER: US/10/301,822
;  CURRENT FILING DATE:  2002-11-21
;  PRIOR APPLICATION NUMBER: US 60/339,971
;  PRIOR FILING DATE: 2001-12-10
;  PRIOR APPLICATION NUMBER: US 60/361,978
;  PRIOR FILING DATE: 2002-03-05
;  PRIOR APPLICATION NUMBER: US 60/381,988
;  PRIOR FILING DATE: 2002-05-20
;  NUMBER OF SEQ ID NOS: 228
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 168
;   LENGTH: 5218
;   TYPE: DNA
;   ORGANISM: Homo Sapiens
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (50)...(1216)
US-10-301-822-168

  Query Match             99.8%;  Score 1510.8;  DB 16;  Length 5218;
  Best Local Similarity   99.9%;  Pred. No. 0;
  Matches 1512;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy          1 ATGCCCGCGGGCTCCTGGGACCCGGCCGGTTACCTGCTCTACTGCCCCTGCATGGGGCGC 60
              ||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        119 ATGCCTGCGGGCTCCTGGGACCCGGCCGGTTACCTGCTCTACTGCCCCTGCATGGGGCGC 178

Qy         61 TTTGGGAACCAGGCCGATCACTTCTTGGGCTCTCTGGCATTTGCAAAGCTGCTAAACCGT 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        179 TTTGGGAACCAGGCCGATCACTTCTTGGGCTCTCTGGCATTTGCAAAGCTGCTAAACCGT 238

Qy        121 ACCTTGGCTGTCCCTCCTTGGATTGAGTACCAGCATCACAAGCCTCCTTTCACCAACCTC 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        239 ACCTTGGCTGTCCCTCCTTGGATTGAGTACCAGCATCACAAGCCTCCTTTCACCAACCTC 298

Qy        181 CATGTGTCCTACCAGAAGTACTTCAAGCTGGAGCCCCTCCAGGCTTACCATCGGGTCATC 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        299 CATGTGTCCTACCAGAAGTACTTCAAGCTGGAGCCCCTCCAGGCTTACCATCGGGTCATC 358

Qy        241 AGCTTGGAGGATTTCATGGAGAAGCTGGCACCCACCCACTGGCCCCCTGAGAAGCGGGTG 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        359 AGCTTGGAGGATTTCATGGAGAAGCTGGCACCCACCCACTGGCCCCCTGAGAAGCGGGTG 418

Qy        301 GCATACTGCTTTGAGGTGGCAGCCCAGCGAAGCCCAGATAAGAAGACGTGCCCCATGAAG 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        419 GCATACTGCTTTGAGGTGGCAGCCCAGCGAAGCCCAGATAAGAAGACGTGCCCCATGAAG 478

Qy        361 GAAGGAAACCCCTTTGGCCCATTCTGGGATCAGTTTCATGTGAGTTTCAACAAGTCGGAG 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        479 GAAGGAAACCCCTTTGGCCCATTCTGGGATCAGTTTCATGTGAGTTTCAACAAGTCGGAG 538

Qy        421 CTTTTTACAGGCATTTCCTTCAGTGCTTCCTACAGAGAACAATGGAGCCAGAGATTTTCT 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        539 CTTTTTACAGGCATTTCCTTCAGTGCTTCCTACAGAGAACAATGGAGCCAGAGATTTTCT 598

Qy        481 CCAAAGGAACATCCGGTGCTTGCCCTGCCAGGAGCCCCAGCCCAGTTCCCCGTCCTAGAA 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        599 CCAAAGGAACATCCGGTGCTTGCCCTGCCAGGAGCCCCAGCCCAGTTCCCCGTCCTAGAG 658

Qy        541 GAACACAGGCCACTACAGAAGTACATGGTATGGTCAGACGAAATGGTGAAGACGGGAGAG 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        659 GAACACAGGCCACTACAGAAGTACATGGTATGGTCAGACGAAATGGTGAAGACGGGAGAG 718

Qy        601 GCCCAGATTCATGCCCACCTTGTCCGGCCCTATGTGGGCATTCATCTGCGCATTGGCTCT 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        719 GCCCAGATTCATGCCCACCTTGTCCGGCCCTATGTGGGCATTCATCTGCGCATTGGCTCT 778

Qy        661 GACTGGAAGAACGCCTGTGCCATGCTGAAGGACGGGACTGCAGGCTCGCACTTCATGGCC 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        779 GACTGGAAGAACGCCTGTGCCATGCTGAAGGACGGGACTGCAGGCTCGCACTTCATGGCC 838

Qy        721 TCTCCGCAGTGTGTGGGCTACAGCCGCAGCACAGCGGCCCCCCTCACGATGACTATGTGC 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        839 TCTCCGCAGTGTGTGGGCTACAGCCGCAGCACAGCGGCCCCCCTCACGATGACTATGTGC 898

Qy        781 CTGCCTGACCTGAAGGAGATCCAGAGGGCTGTGAAGCTCTGGGTGAGGTCGCTGGATGCC 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        899 CTGCCTGACCTGAAGGAGATCCAGAGGGCTGTGAAGCTCTGGGTGAGGTCGCTGGATGCC 958

Qy        841 CAGTCGGTCTACGTTGCTACTGATTCCGAGAGTTATGTGCCTGAGCTCCAACAGCTCTTC 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        959 CAGTCGGTCTACGTTGCTACTGATTCCGAGAGTTATGTGCCTGAGCTCCAACAGCTCTTC 1018

Qy        901 AAAGGGAAGGTGAAGGTGGTGAGCCTGAAGCCTGAGGTGGCCCAGGTCGACCTGTACATC 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1019 AAAGGGAAGGTGAAGGTGGTGAGCCTGAAGCCTGAGGTGGCCCAGGTCGACCTGTACATC 1078

Qy        961 CTCGGCCAAGCCGACCACTTTATTGGCAACTGTGTCTCCTCCTTCACTGCCTTTGTGAAG 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1079 CTCGGCCAAGCCGACCACTTTATTGGCAACTGTGTCTCCTCCTTCACTGCCTTTGTGAAG 1138

Qy       1021 CGGGAGCGGGACCTCCAGGGGAGGCCGTCTTCTTTCTTCGGCATGGACAGGCCCCCTAAG 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1139 CGGGAGCGGGACCTCCAGGGGAGGCCGTCTTCTTTCTTCGGCATGGACAGGCCCCCTAAG 1198

Qy       1081 CTGCGGGACGAGTTCTGATTCTGGCCGGAGCACCAGACCCTCTGATCCTGGAGGGACCAG 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1199 CTGCGGGACGAGTTCTGATTCTGGCCGGAGCACCAGACCCTCTGATCCTGGAGGGACCAG 1258

Qy       1141 AGTCTGAGCTGGTCCTTCCAGCCAGGCCTGGCAGCCAGAGGTGCTCCGGGATTGCAAACT 1200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1259 AGTCTGAGCTGGTCCTTCCAGCCAGGCCTGGCAGCCAGAGGTGCTCCGGGATTGCAAACT 1318

Qy       1201 CCTCTTCTCACCTGCCAAAGATGGAGAAGAGTGCCAGGGACCCCTCAAGGAGGGAGACGC 1260
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1319 CCTCTTCTCACCTGCCAAAGATGGAGAAGAGTGCCAGGGACCCCTCAAGGAGGGAGACGC 1378

Qy       1261 TCCATATCCCAGGGCATAGGACTTGCAGGTTCCTAGGAGCAGGAGCATCTCCCATCGCAC 1320
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1379 TCCATATCCCAGGGCATAGGACTTGCAGGTTCCTAGGAGCAGGAGCATCTCCCATCGCAC 1438

Qy       1321 GTGCTTTCTGCTCTTCTGGGAATTTCTCACACTGGCAAAGCAGTCCAGCCTCCGTCTTCT 1380
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1439 GTGCTTTCTGCTCTTCTGGGAATTTCTCACACTGGCAAAGCAGTCCAGCCTCCGTCTTCT 1498

Qy       1381 GGTCCACTCTGCTCTGAGCAGCCTGGGATGCTGAACTCTTCAGAGAGATTTTTTTATAGA 1440
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1499 GGTCCACTCTGCTCTGAGCAGCCTGGGATGCTGAACTCTTCAGAGAGATTTTTTTATAGA 1558

Qy       1441 GAGATTTCTATAATTTTGATACAAGGTCATGACTATCCTAGAACTCTCTGTGGTTTTTGA 1500
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1559 GAGATTTCTATAATTTTGATACAAGGTCATGACTATCCTAGAACTCTCTGTGGTTTTTGA 1618

Qy       1501 AAATCATTGAATTC 1514
              ||||||||||||||
Db       1619 AAATCATTGAATTC 1632


RESULT 4
US-10-956-157-4752
; Sequence 4752, Application US/10956157
; Publication No. US20050118625A1
; GENERAL INFORMATION:
;  APPLICANT: Wyeth
;  APPLICANT:  Mounts, William
;  TITLE OF INVENTION: NUCLEIC ACID ARRAYS FOR DETECTING GENE EXPRESSION ASSOCIATED WITH
;  TITLE OF INVENTION:  HUMAN OSTEOARTHRITIS AND HUMAN PROTEASES
;  FILE REFERENCE: 031896-043000 (AM 101081)
;  CURRENT APPLICATION NUMBER: US/10/956,157
;  CURRENT FILING DATE:  2004-10-04
;  NUMBER OF SEQ ID NOS: 319805
```

```
;  SOFTWARE: PatentIn version 3.2
; SEQ ID NO 4752
;   LENGTH: 5218
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-10-956-157-4752

  Query Match             99.8%;  Score 1510.8;  DB 22;  Length 5218;
  Best Local Similarity   99.9%;  Pred. No. 0;
  Matches 1512;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy          1 ATGCCCGCGGGCTCCTGGGACCCGGCCGGTTACCTGCTCTACTGCCCCTGCATGGGGCGC 60
              ||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        119 ATGCCTGCGGGCTCCTGGGACCCGGCCGGTTACCTGCTCTACTGCCCCTGCATGGGGCGC 178

Qy         61 TTTGGGAACCAGGCCGATCACTTCTTGGGCTCTCTGGCATTTGCAAAGCTGCTAAACCGT 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        179 TTTGGGAACCAGGCCGATCACTTCTTGGGCTCTCTGGCATTTGCAAAGCTGCTAAACCGT 238

Qy        121 ACCTTGGCTGTCCCTCCTTGGATTGAGTACCAGCATCACAAGCCTCCTTTCACCAACCTC 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        239 ACCTTGGCTGTCCCTCCTTGGATTGAGTACCAGCATCACAAGCCTCCTTTCACCAACCTC 298

Qy        181 CATGTGTCCTACCAGAAGTACTTCAAGCTGGAGCCCCTCCAGGCTTACCATCGGGTCATC 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        299 CATGTGTCCTACCAGAAGTACTTCAAGCTGGAGCCCCTCCAGGCTTACCATCGGGTCATC 358

Qy        241 AGCTTGGAGGATTTCATGGAGAAGCTGGCACCCACCCACTGGCCCCCTGAGAAGCGGGTG 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        359 AGCTTGGAGGATTTCATGGAGAAGCTGGCACCCACCCACTGGCCCCCTGAGAAGCGGGTG 418

Qy        301 GCATACTGCTTTGAGGTGGCAGCCCAGCGAAGCCCAGATAAGAAGACGTGCCCCATGAAG 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        419 GCATACTGCTTTGAGGTGGCAGCCCAGCGAAGCCCAGATAAGAAGACGTGCCCCATGAAG 478

Qy        361 GAAGGAAACCCCTTTGGCCCATTCTGGGATCAGTTTCATGTGAGTTTCAACAAGTCGGAG 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        479 GAAGGAAACCCCTTTGGCCCATTCTGGGATCAGTTTCATGTGAGTTTCAACAAGTCGGAG 538

Qy        421 CTTTTTACAGGCATTTCCTTCAGTGCTTCCTACAGAGAACAATGGAGCCAGAGATTTTCT 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        539 CTTTTTACAGGCATTTCCTTCAGTGCTTCCTACAGAGAACAATGGAGCCAGAGATTTTCT 598

Qy        481 CCAAAGGAACATCCGGTGCTTGCCCTGCCAGGAGCCCCAGCCCAGTTCCCCGTCCTAGAA 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        599 CCAAAGGAACATCCGGTGCTTGCCCTGCCAGGAGCCCCAGCCCAGTTCCCCGTCCTAGAG 658

Qy        541 GAACACAGGCCACTACAGAAGTACATGGTATGGTCAGACGAAATGGTGAAGACGGGAGAG 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        659 GAACACAGGCCACTACAGAAGTACATGGTATGGTCAGACGAAATGGTGAAGACGGGAGAG 718

Qy        601 GCCCAGATTCATGCCCACCTTGTCCGGCCCTATGTGGGCATTCATCTGCGCATTGGCTCT 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        719 GCCCAGATTCATGCCCACCTTGTCCGGCCCTATGTGGGCATTCATCTGCGCATTGGCTCT 778

Qy        661 GACTGGAAGAACGCCTGTGCCATGCTGAAGGACGGGACTGCAGGCTCGCACTTCATGGCC 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        779 GACTGGAAGAACGCCTGTGCCATGCTGAAGGACGGGACTGCAGGCTCGCACTTCATGGCC 838

Qy        721 TCTCCGCAGTGTGTGGGCTACAGCCGCAGCACAGCGGCCCCCCTCACGATGACTATGTGC 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        839 TCTCCGCAGTGTGTGGGCTACAGCCGCAGCACAGCGGCCCCCCTCACGATGACTATGTGC 898

Qy        781 CTGCCTGACCTGAAGGAGATCCAGAGGGCTGTGAAGCTCTGGGTGAGGTCGCTGGATGCC 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        899 CTGCCTGACCTGAAGGAGATCCAGAGGGCTGTGAAGCTCTGGGTGAGGTCGCTGGATGCC 958

Qy        841 CAGTCGGTCTACGTTGCTACTGATTCCGAGAGTTATGTGCCTGAGCTCCAACAGCTCTTC 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        959 CAGTCGGTCTACGTTGCTACTGATTCCGAGAGTTATGTGCCTGAGCTCCAACAGCTCTTC 1018

Qy        901 AAAGGGAAGGTGAAGGTGGTGAGCCTGAAGCCTGAGGTGGCCCAGGTCGACCTGTACATC 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

Db       1019 AAAGGGAAGGTGAAGGTGGTGAGCCTGAAGCCTGAGGTGGCCCAGGTCGACCTGTACATC 1078

Qy        961 CTCGGCCAAGCCGACCACTTTATTGGCAACTGTGTCTCCTCCTTCACTGCCTTTGTGAAG 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1079 CTCGGCCAAGCCGACCACTTTATTGGCAACTGTGTCTCCTCCTTCACTGCCTTTGTGAAG 1138

Qy       1021 CGGGAGCGGGACCTCCAGGGGAGGCCGTCTTCTTTCTTCGGCATGGACAGGCCCCCTAAG 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1139 CGGGAGCGGGACCTCCAGGGGAGGCCGTCTTCTTTCTTCGGCATGGACAGGCCCCCTAAG 1198

Qy       1081 CTGCGGGACGAGTTCTGATTCTGGCCGGAGCACCAGACCCTCTGATCCTGGAGGGACCAG 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1199 CTGCGGGACGAGTTCTGATTCTGGCCGGAGCACCAGACCCTCTGATCCTGGAGGGACCAG 1258

Qy       1141 AGTCTGAGCTGGTCCTTCCAGCCAGGCCTGGCAGCCAGAGGTGCTCCGGGATTGCAAACT 1200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1259 AGTCTGAGCTGGTCCTTCCAGCCAGGCCTGGCAGCCAGAGGTGCTCCGGGATTGCAAACT 1318

Qy       1201 CCTCTTCTCACCTGCCAAAGATGGAGAAGAGTGCCAGGGACCCCTCAAGGAGGGAGACGC 1260
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1319 CCTCTTCTCACCTGCCAAAGATGGAGAAGAGTGCCAGGGACCCCTCAAGGAGGGAGACGC 1378

Qy       1261 TCCATATCCCAGGGCATAGGACTTGCAGGTTCCTAGGAGCAGGAGCATCTCCCATCGCAC 1320
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1379 TCCATATCCCAGGGCATAGGACTTGCAGGTTCCTAGGAGCAGGAGCATCTCCCATCGCAC 1438

Qy       1321 GTGCTTTCTGCTCTTCTGGGAATTTCTCACACTGGCAAAGCAGTCCAGCCTCCGTCTTCT 1380
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1439 GTGCTTTCTGCTCTTCTGGGAATTTCTCACACTGGCAAAGCAGTCCAGCCTCCGTCTTCT 1498

Qy       1381 GGTCCACTCTGCTCTGAGCAGCCTGGGATGCTGAACTCTTCAGAGAGATTTTTTTATAGA 1440
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1499 GGTCCACTCTGCTCTGAGCAGCCTGGGATGCTGAACTCTTCAGAGAGATTTTTTTATAGA 1558

Qy       1441 GAGATTTCTATAATTTTGATACAAGGTCATGACTATCCTAGAACTCTCTGTGGTTTTTGA 1500
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1559 GAGATTTCTATAATTTTGATACAAGGTCATGACTATCCTAGAACTCTCTGTGGTTTTTGA 1618

Qy       1501 AAATCATTGAATTC 1514
              ||||||||||||||
Db       1619 AAATCATTGAATTC 1632


RESULT 5
US-09-774-954-7
; Sequence 7, Application US/09774954
; Publication No. US20040241645A1
;  GENERAL INFORMATION:
;       APPLICANT: Yang Wang, Michael W. Spellman
;       TITLE OF INVENTION: O-Fucosyltransferase
;       NUMBER OF SEQUENCES: 17
;       CORRESPONDENCE ADDRESS:
;            ADDRESSEE: Genentech, Inc.
;            STREET: 1 DNA Way
;            CITY: South San Francisco
;            STATE: California
;            COUNTRY: USA
;            ZIP: 94080
;       COMPUTER READABLE FORM:
;            MEDIUM TYPE: 3.5 inch, 1.44 Mb floppy disk
;            COMPUTER: IBM PC compatible
;            OPERATING SYSTEM: PC-DOS/MS-DOS
;            SOFTWARE: WinPatin (Genentech)
;       CURRENT APPLICATION DATA:
;            APPLICATION NUMBER: US/09/774,954
;            FILING DATE: 30-Jan-2001
;            CLASSIFICATION: <Unknown>
;       PRIOR APPLICATION DATA:
;            APPLICATION NUMBER: US/08/978,741
;            FILING DATE: 26-NOV-1997
;            APPLICATION NUMBER: 08/792,498
;            FILING DATE: 31-JAN-1997
;       ATTORNEY/AGENT INFORMATION:
```

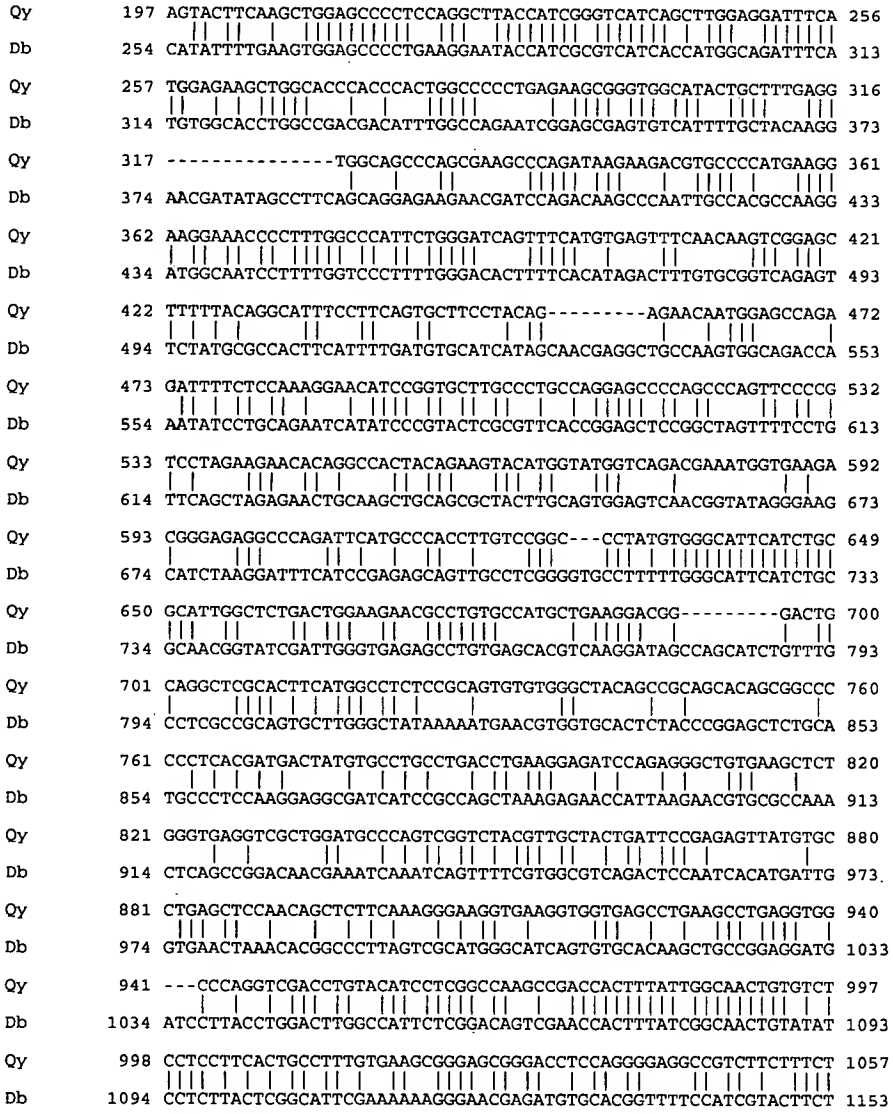
```
;             NAME: Svoboda, Craig G.
;             REGISTRATION NUMBER: 39,044
;             REFERENCE/DOCKET NUMBER: P1041P1
;       TELECOMMUNICATION INFORMATION:
;             TELEPHONE: 650/225-1489
;             TELEFAX: 650/952-9881
;   INFORMATION FOR SEQ ID NO: 7:
;       SEQUENCE CHARACTERISTICS:
;             LENGTH: 5009 base pairs
;             TYPE: Nucleic Acid
;             STRANDEDNESS: Single
;             TOPOLOGY: Linear
;       SEQUENCE DESCRIPTION: SEQ ID NO: 7:
US-09-774-954-7

  Query Match             95.6%;  Score 1447.4;  DB 11;  Length 5009;
  Best Local Similarity   99.9%;  Pred. No. 0;
  Matches 1448;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy         66 GAACCAGGCCGATCACTTCTTGGGCTCTCTGGCATTTGCAAAGCTGCTAAACCGTACCTT 125
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 GAACCAGGCCGATCACTTCTTGGGCTCTCTGGCATTTGCAAAGCTGCTAAACCGTACCTT 60

Qy        126 GGCTGTCCCTCCTTGGATTGAGTACCAGCATCACAAGCCTCCTTTCACCAACCTCCATGT 185
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GGCTGTCCCTCCTTGGATTGAGTACCAGCATCACAAGCCTCCTTTCACCAACCTCCATGT 120

Qy        186 GTCCTACCAGAAGTACTTCAAGCTGGAGCCCCTCCAGGCTTACCATCGGGTCATCAGCTT 245
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 GTCCTACCAGAAGTACTTCAAGCTGGAGCCCCTCCAGGCTTACCATCGGGTCATCAGCTT 180

Qy        246 GGAGGATTTCATGGAGAAGCTGGCACCCACCCACTGGCCCCCTGAGAAGCGGGTGGCATA 305
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 GGAGGATTTCATGGAGAAGCTGGCACCCACCCACTGGCCCCCTGAGAAGCGGGTGGCATA 240

Qy        306 CTGCTTTGAGGTGGCAGCCCAGCGAAGCCCAGATAAGAAGACGTGCCCCATGAAGGAAGG 365
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 CTGCTTTGAGGTGGCAGCCCAGCGAAGCCCAGATAAGAAGACGTGCCCCATGAAGGAAGG 300

Qy        366 AAACCCCTTTGGCCCATTCTGGGATCAGTTTCATGTGAGTTTCAACAAGTCGGAGCTTTT 425
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 AAACCCCTTTGGCCCATTCTGGGATCAGTTTCATGTGAGTTTCAACAAGTCGGAGCTTTT 360

Qy        426 TACAGGCATTTCCTTCAGTGCTTCCTACAGAGAACAATGGAGCCAGAGATTTTCTCCAAA 485
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 TACAGGCATTTCCTTCAGTGCTTCCTACAGAGAACAATGGAGCCAGAGATTTTCTCCAAA 420

Qy        486 GGAACATCCGGTGCTTGCCCTGCCAGGAGCCCCAGCCCAGTTCCCCGTCCTAGAAGAACA 545
              |||||||||||||||||||||||||||||||||||||||||||||||||||| |||||
Db        421 GGAACATCCGGTGCTTGCCCTGCCAGGAGCCCCAGCCCAGTTCCCCGTCCTAGAGGAACA 480

Qy        546 CAGGCCACTACAGAAGTACATGGTATGGTCAGACGAAATGGTGAAGACGGGAGAGGCCCA 605
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 CAGGCCACTACAGAAGTACATGGTATGGTCAGACGAAATGGTGAAGACGGGAGAGGCCCA 540

Qy        606 GATTCATGCCCACCTTGTCCGGCCCTATGTGGGCATTCATCTGCGCATTGGCTCTGACTG 665
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 GATTCATGCCCACCTTGTCCGGCCCTATGTGGGCATTCATCTGCGCATTGGCTCTGACTG 600

Qy        666 GAAGAACGCCTGTGCCATGCTGAAGGACGGGACTGCAGGCTCGCACTTCATGGCCTCTCC 725
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 GAAGAACGCCTGTGCCATGCTGAAGGACGGGACTGCAGGCTCGCACTTCATGGCCTCTCC 660

Qy        726 GCAGTGTGTGGGCTACAGCCGCAGCACAGCGGCCCCCCTCACGATGACTATGTGCCTGCC 785
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 GCAGTGTGTGGGCTACAGCCGCAGCACAGCGGCCCCCCTCACGATGACTATGTGCCTGCC 720

Qy        786 TGACCTGAAGGAGATCCAGAGGGCTGTGAAGCTCTGGGTGAGGTCGCTGGATGCCCAGTC 845
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 TGACCTGAAGGAGATCCAGAGGGCTGTGAAGCTCTGGGTGAGGTCGCTGGATGCCCAGTC 780

Qy        846 GGTCTACGTTGCTACTGATTCCGAGAGTTATGTGCCTGAGCTCCAACAGCTCTTCAAAGG 905
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

Db        781 GGTCTACGTTGCTACTGATTCCGAGAGTTATGTGCCTGAGCTCCAACAGCTCTTCAAAGG 840

Qy        906 GAAGGTGAAGGTGGTGAGCCTGAAGCCTGAGGTGGCCCAGGTCGACCTGTACATCCTCGG 965
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 GAAGGTGAAGGTGGTGAGCCTGAAGCCTGAGGTGGCCCAGGTCGACCTGTACATCCTCGG 900

Qy        966 CCAAGCCGACCACTTTATTGGCAACTGTGTCTCCTCCTTCACTGCCTTTGTGAAGCGGGA 1025
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 CCAAGCCGACCACTTTATTGGCAACTGTGTCTCCTCCTTCACTGCCTTTGTGAAGCGGGA 960

Qy       1026 GCGGGACCTCCAGGGGAGGCCGTCTTCTTTCTTCGGCATGGACAGGCCCCCTAAGCTGCG 1085
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 GCGGGACCTCCAGGGGAGGCCGTCTTCTTTCTTCGGCATGGACAGGCCCCCTAAGCTGCG 1020

Qy       1086 GGACGAGTTCTGATTCTGGCCGGAGCACCAGACCCTCTGATCCTGGAGGGACCAGAGTCT 1145
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1021 GGACGAGTTCTGATTCTGGCCGGAGCACCAGACCCTCTGATCCTGGAGGGACCAGAGTCT 1080

Qy       1146 GAGCTGGTCCTTCCAGCCAGGCCTGGCAGCCAGAGGTGCTCCGGGATTGCAAACTCCTCT 1205
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1081 GAGCTGGTCCTTCCAGCCAGGCCTGGCAGCCAGAGGTGCTCCGGGATTGCAAACTCCTCT 1140

Qy       1206 TCTCACCTGCCAAAGATGGAGAAGAGTGCCAGGGACCCCTCAAGGAGGGAGACGCTCCAT 1265
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1141 TCTCACCTGCCAAAGATGGAGAAGAGTGCCAGGGACCCCTCAAGGAGGGAGACGCTCCAT 1200

Qy       1266 ATCCCAGGGCATAGGACTTGCAGGTTCCTAGGAGCAGGAGCATCTCCCATCGCACGTGCT 1325
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1201 ATCCCAGGGCATAGGACTTGCAGGTTCCTAGGAGCAGGAGCATCTCCCATCGCACGTGCT 1260

Qy       1326 TTCTGCTCTTCTGGGAATTTCTCACACTGGCAAAGCAGTCCAGCCTCCGTCTTCTGGTCC 1385
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1261 TTCTGCTCTTCTGGGAATTTCTCACACTGGCAAAGCAGTCCAGCCTCCGTCTTCTGGTCC 1320

Qy       1386 ACTCTGCTCTGAGCAGCCTGGGATGCTGAACTCTTCAGAGAGATTTTTTTATAGAGAGAT 1445
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1321 ACTCTGCTCTGAGCAGCCTGGGATGCTGAACTCTTCAGAGAGATTTTTTTATAGAGAGAT 1380

Qy       1446 TTCTATAATTTTGATACAAGGTCATGACTATCCTAGAACTCTCTGTGGTTTTTGAAAATC 1505
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1381 TTCTATAATTTTGATACAAGGTCATGACTATCCTAGAACTCTCTGTGGTTTTTGAAAATC 1440

Qy       1506 ATTGAATTC 1514
              |||||||||
Db       1441 ATTGAATTC 1449


RESULT 6
US-09-774-954-4
; Sequence 4, Application US/09774954
; Publication No. US20040241645A1
;   GENERAL INFORMATION:
;        APPLICANT: Yang Wang, Michael W. Spellman
;        TITLE OF INVENTION: O-Fucosyltransferase
;        NUMBER OF SEQUENCES: 17
;        CORRESPONDENCE ADDRESS:
;             ADDRESSEE: Genentech, Inc.
;             STREET: 1 DNA Way
;             CITY: South San Francisco
;             STATE: California
;             COUNTRY: USA
;             ZIP: 94080
;        COMPUTER READABLE FORM:
;             MEDIUM TYPE: 3.5 inch, 1.44 Mb floppy disk
;             COMPUTER: IBM PC compatible
;             OPERATING SYSTEM: PC-DOS/MS-DOS
;             SOFTWARE: WinPatin (Genentech)
;        CURRENT APPLICATION DATA:
;             APPLICATION NUMBER: US/09/774,954
;             FILING DATE: 30-Jan-2001
;             CLASSIFICATION: <Unknown>
;        PRIOR APPLICATION DATA:
;             APPLICATION NUMBER: US/08/978,741
```

```
;            FILING DATE: 26-NOV-1997
;            APPLICATION NUMBER: 08/792,498
;            FILING DATE: 31-JAN-1997
;       ATTORNEY/AGENT INFORMATION:
;            NAME: Svoboda, Craig G.
;            REGISTRATION NUMBER: 39,044
;            REFERENCE/DOCKET NUMBER: P1041P1
;       TELECOMMUNICATION INFORMATION:
;            TELEPHONE: 650/225-1489
;            TELEFAX: 650/952-9881
;   INFORMATION FOR SEQ ID NO: 4:
;       SEQUENCE CHARACTERISTICS:
;            LENGTH: 1300 base pairs
;            TYPE: Nucleic Acid
;            STRANDEDNESS: Single
;            TOPOLOGY: Linear
;       SEQUENCE DESCRIPTION: SEQ ID NO: 4:
US-09-774-954-4

  Query Match             76.8%;  Score 1163.4;  DB 11;  Length 1300;
  Best Local Similarity   99.9%;  Pred. No. 0;
  Matches 1164;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 ATGCCCGCGGGCTCCTGGGACCCGGCCGGTTACCTGCTCTACTGCCCCTGCATGGGGCGC 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        136 ATGCCCGCGGGCTCCTGGGACCCGGCCGGTTACCTGCTCTACTGCCCCTGCATGGGGCGC 195

Qy         61 TTTGGGAACCAGGCCGATCACTTCTTGGGCTCTCTGGCATTTGCAAAGCTGCTAAACCGT 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        196 TTTGGGAACCAGGCCGATCACTTCTTGGGCTCTCTGGCATTTGCAAAGCTGCTAAACCGT 255

Qy        121 ACCTTGGCTGTCCCTCCTTGGATTGAGTACCAGCATCACAAGCCTCCTTTCACCAACCTC 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        256 ACCTTGGCTGTCCCTCCTTGGATTGAGTACCAGCATCACAAGCCTCCTTTCACCAACCTC 315

Qy        181 CATGTGTCCTACCAGAAGTACTTCAAGCTGGAGCCCCTCCAGGCTTACCATCGGGTCATC 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        316 CATGTGTCCTACCAGAAGTACTTCAAGCTGGAGCCCCTCCAGGCTTACCATCGGGTCATC 375

Qy        241 AGCTTGGAGGATTTCATGGAGAAGCTGGCACCCACCCACTGGCCCCCTGAGAAGCGGGTG 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        376 AGCTTGGAGGATTTCATGGAGAAGCTGGCACCCACCCACTGGCCCCCTGAGAAGCGGGTG 435

Qy        301 GCATACTGCTTTGAGGTGGCAGCCCAGCGAAGCCCAGATAAGAAGACGTGCCCCATGAAG 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        436 GCATACTGCTTTGAGGTGGCAGCCCAGCGAAGCCCAGATAAGAAGACGTGCCCCATGAAG 495

Qy        361 GAAGGAAACCCCTTTGGCCCATTCTGGGATCAGTTTCATGTGAGTTTCAACAAGTCGGAG 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        496 GAAGGAAACCCCTTTGGCCCATTCTGGGATCAGTTTCATGTGAGTTTCAACAAGTCGGAG 555

Qy        421 CTTTTTACAGGCATTTCCTTCAGTGCTTCCTACAGAGAACAATGGAGCCAGAGATTTTCT 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        556 CTTTTTACAGGCATTTCCTTCAGTGCTTCCTACAGAGAACAATGGAGCCAGAGATTTTCT 615

Qy        481 CCAAAGGAACATCCGGTGCTTGCCCTGCCAGGAGCCCCAGCCCAGTTCCCCGTCCTAGAA 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        616 CCAAAGGAACATCCGGTGCTTGCCCTGCCAGGAGCCCCAGCCCAGTTCCCCGTCCTAGAG 675

Qy        541 GAACACAGGCCACTACAGAAGTACATGGTATGGTCAGACGAAATGGTGAAGACGGGAGAG 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        676 GAACACAGGCCACTACAGAAGTACATGGTATGGTCAGACGAAATGGTGAAGACGGGAGAG 735

Qy        601 GCCCAGATTCATGCCCACCTTGTCCGGCCCTATGTGGGCATTCATCTGCGCATTGGCTCT 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        736 GCCCAGATTCATGCCCACCTTGTCCGGCCCTATGTGGGCATTCATCTGCGCATTGGCTCT 795

Qy        661 GACTGGAAGAACGCCTGTGCCATGCTGAAGGACGGGACTGCAGGCTCGCACTTCATGGCC 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        796 GACTGGAAGAACGCCTGTGCCATGCTGAAGGACGGGACTGCAGGCTCGCACTTCATGGCC 855

Qy        721 TCTCCGCAGTGTGTGGGCTACAGCCGCAGCACAGCGGCCCCCCTCACGATGACTATGTGC 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        856 TCTCCGCAGTGTGTGGGCTACAGCCGCAGCACAGCGGCCCCCCTCACGATGACTATGTGC 915

Qy        781 CTGCCTGACCTGAAGGAGATCCAGAGGGCTGTGAAGCTCTGGGTGAGGTCGCTGGATGCC 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        916 CTGCCTGACCTGAAGGAGATCCAGAGGGCTGTGAAGCTCTGGGTGAGGTCGCTGGATGCC 975

Qy        841 CAGTCGGTCTACGTTGCTACTGATTCCGAGAGTTATGTGCCTGAGCTCCAACAGCTCTTC 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        976 CAGTCGGTCTACGTTGCTACTGATTCCGAGAGTTATGTGCCTGAGCTCCAACAGCTCTTC 1035

Qy        901 AAAGGGAAGGTGAAGGTGGTGAGCCTGAAGCCTGAGGTGGCCCAGGTCGACCTGTACATC 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1036 AAAGGGAAGGTGAAGGTGGTGAGCCTGAAGCCTGAGGTGGCCCAGGTCGACCTGTACATC 1095

Qy        961 CTCGGCCAAGCCGACCACTTTATTGGCAACTGTGTCTCCTCCTTCACTGCCTTTGTGAAG 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1096 CTCGGCCAAGCCGACCACTTTATTGGCAACTGTGTCTCCTCCTTCACTGCCTTTGTGAAG 1155

Qy       1021 CGGGAGCGGGACCTCCAGGGGAGGCCGTCTTCTTTCTTCGGCATGGACAGGCCCCCTAAG 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1156 CGGGAGCGGGACCTCCAGGGGAGGCCGTCTTCTTTCTTCGGCATGGACAGGCCCCCTAAG 1215

Qy       1081 CTGCGGGACGAGTTCTGATTCTGGCCGGAGCACCAGACCCTCTGATCCTGGAGGGACCAG 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1216 CTGCGGGACGAGTTCTGATTCTGGCCGGAGCACCAGACCCTCTGATCCTGGAGGGACCAG 1275

Qy       1141 AGTCTGAGCTGGTCCTTCCAGCCAG 1165
              |||||||||||||||||||||||||
Db       1276 AGTCTGAGCTGGTCCTTCCAGCCAG 1300


RESULT 7
US-09-774-954-16
; Sequence 16, Application US/09774954
; Publication No. US20040241645A1
;   GENERAL INFORMATION:
;       APPLICANT: Yang Wang, Michael W. Spellman
;       TITLE OF INVENTION: O-Fucosyltransferase
;       NUMBER OF SEQUENCES: 17
;       CORRESPONDENCE ADDRESS:
;            ADDRESSEE: Genentech, Inc.
;            STREET: 1 DNA Way
;            CITY: South San Francisco
;            STATE: California
;            COUNTRY: USA
;            ZIP: 94080
;       COMPUTER READABLE FORM:
;            MEDIUM TYPE: 3.5 inch, 1.44 Mb floppy disk
;            COMPUTER: IBM PC compatible
;            OPERATING SYSTEM: PC-DOS/MS-DOS
;            SOFTWARE: WinPatin (Genentech)
;       CURRENT APPLICATION DATA:
;            APPLICATION NUMBER: US/09/774,954
;            FILING DATE: 30-Jan-2001
;            CLASSIFICATION: <Unknown>
;       PRIOR APPLICATION DATA:
;            APPLICATION NUMBER: US/08/978,741
;            FILING DATE: 26-NOV-1997
;            APPLICATION NUMBER: 08/792,498
;            FILING DATE: 31-JAN-1997
;       ATTORNEY/AGENT INFORMATION:
;            NAME: Svoboda, Craig G.
;            REGISTRATION NUMBER: 39,044
;            REFERENCE/DOCKET NUMBER: P1041P1
;       TELECOMMUNICATION INFORMATION:
;            TELEPHONE: 650/225-1489
;            TELEFAX: 650/952-9881
;   INFORMATION FOR SEQ ID NO: 16:
;       SEQUENCE CHARACTERISTICS:
;            LENGTH: 1100 base pairs
;            TYPE: Nucleic Acid
;            STRANDEDNESS: Single
;            TOPOLOGY: Linear
```

```
;        SEQUENCE DESCRIPTION: SEQ ID NO: 16:
US-09-774-954-16

  Query Match             72.7%;  Score 1100;  DB 11;  Length 1100;
  Best Local Similarity   100.0%;  Pred. No. 0;
  Matches 1100;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 ATGCCCGCGGGCTCCTGGGACCCGGCCGGTTACCTGCTCTACTGCCCCTGCATGGGGCGC 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGCCCGCGGGCTCCTGGGACCCGGCCGGTTACCTGCTCTACTGCCCCTGCATGGGGCGC 60

Qy         61 TTTGGGAACCAGGCCGATCACTTCTTGGGCTCTCTGGCATTTGCAAAGCTGCTAAACCGT 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 TTTGGGAACCAGGCCGATCACTTCTTGGGCTCTCTGGCATTTGCAAAGCTGCTAAACCGT 120

Qy        121 ACCTTGGCTGTCCCTCCTTGGATTGAGTACCAGCATCACAAGCCTCCTTTCACCAACCTC 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 ACCTTGGCTGTCCCTCCTTGGATTGAGTACCAGCATCACAAGCCTCCTTTCACCAACCTC 180

Qy        181 CATGTGTCCTACCAGAAGTACTTCAAGCTGGAGCCCCTCCAGGCTTACCATCGGGTCATC 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 CATGTGTCCTACCAGAAGTACTTCAAGCTGGAGCCCCTCCAGGCTTACCATCGGGTCATC 240

Qy        241 AGCTTGGAGGATTTCATGGAGAAGCTGGCACCCACCCACTGGCCCCCTGAGAAGCGGGTG 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 AGCTTGGAGGATTTCATGGAGAAGCTGGCACCCACCCACTGGCCCCCTGAGAAGCGGGTG 300

Qy        301 GCATACTGCTTTGAGGTGGCAGCCCAGCGAAGCCCAGATAAGAAGACGTGCCCCATGAAG 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 GCATACTGCTTTGAGGTGGCAGCCCAGCGAAGCCCAGATAAGAAGACGTGCCCCATGAAG 360

Qy        361 GAAGGAAACCCCTTTGGCCCATTCTGGGATCAGTTTCATGTGAGTTTCAACAAGTCGGAG 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 GAAGGAAACCCCTTTGGCCCATTCTGGGATCAGTTTCATGTGAGTTTCAACAAGTCGGAG 420

Qy        421 CTTTTTACAGGCATTTCCTTCAGTGCTTCCTACAGAGAACAATGGAGCCAGAGATTTTCT 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 CTTTTTACAGGCATTTCCTTCAGTGCTTCCTACAGAGAACAATGGAGCCAGAGATTTTCT 480

Qy        481 CCAAAGGAACATCCGGTGCTTGCCCTGCCAGGAGCCCCAGCCCAGTTCCCCGTCCTAGAA 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 CCAAAGGAACATCCGGTGCTTGCCCTGCCAGGAGCCCCAGCCCAGTTCCCCGTCCTAGAA 540

Qy        541 GAACACAGGCCACTACAGAAGTACATGGTATGGTCAGACGAAATGGTGAAGACGGGAGAG 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 GAACACAGGCCACTACAGAAGTACATGGTATGGTCAGACGAAATGGTGAAGACGGGAGAG 600

Qy        601 GCCCAGATTCATGCCCACCTTGTCCGGCCCTATGTGGGCATTCATCTGCGCATTGGCTCT 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 GCCCAGATTCATGCCCACCTTGTCCGGCCCTATGTGGGCATTCATCTGCGCATTGGCTCT 660

Qy        661 GACTGGAAGAACGCCTGTGCCATGCTGAAGGACGGGACTGCAGGCTCGCACTTCATGGCC 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 GACTGGAAGAACGCCTGTGCCATGCTGAAGGACGGGACTGCAGGCTCGCACTTCATGGCC 720

Qy        721 TCTCCGCAGTGTGTGGGCTACAGCCGCAGCACAGCGGCCCCCCTCACGATGACTATGTGC 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 TCTCCGCAGTGTGTGGGCTACAGCCGCAGCACAGCGGCCCCCCTCACGATGACTATGTGC 780

Qy        781 CTGCCTGACCTGAAGGAGATCCAGAGGGCTGTGAAGCTCTGGGTGAGGTCGCTGGATGCC 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 CTGCCTGACCTGAAGGAGATCCAGAGGGCTGTGAAGCTCTGGGTGAGGTCGCTGGATGCC 840

Qy        841 CAGTCGGTCTACGTTGCTACTGATTCCGAGAGTTATGTGCCTGAGCTCCAACAGCTCTTC 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 CAGTCGGTCTACGTTGCTACTGATTCCGAGAGTTATGTGCCTGAGCTCCAACAGCTCTTC 900

Qy        901 AAAGGGAAGGTGAAGGTGGTGAGCCTGAAGCCTGAGGTGGCCCAGGTCGACCTGTACATC 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 AAAGGGAAGGTGAAGGTGGTGAGCCTGAAGCCTGAGGTGGCCCAGGTCGACCTGTACATC 960

Qy        961 CTCGGCCAAGCCGACCACTTTATTGGCAACTGTGTCTCCTCCTTCACTGCCTTTGTGAAG 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 CTCGGCCAAGCCGACCACTTTATTGGCAACTGTGTCTCCTCCTTCACTGCCTTTGTGAAG 1020

Qy       1021 CGGGAGCGGGACCTCCAGGGGAGGCCGTCTTCTTTCTTCGGCATGGACAGGCCCCCTAAG 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1021 CGGGAGCGGGACCTCCAGGGGAGGCCGTCTTCTTTCTTCGGCATGGACAGGCCCCCTAAG 1080

Qy       1081 CTGCGGGACGAGTTCTGATT 1100
              ||||||||||||||||||||
Db       1081 CTGCGGGACGAGTTCTGATT 1100


RESULT 8
US-09-764-891-6032
; Sequence 6032, Application US/09764891
; Publication No. US20030077808A1
; GENERAL INFORMATION:
;  APPLICANT: Rosen et al.
;  TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
;  FILE REFERENCE: PC006
;  CURRENT APPLICATION NUMBER: US/09/764,891
;  CURRENT FILING DATE:  2001-01-17
;  Prior application data removed - consult PALM or file wrapper
;  NUMBER OF SEQ ID NOS: 10231
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 6032
;   LENGTH: 10331
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-09-764-891-6032

  Query Match             40.2%;  Score 608.6;  DB 10;  Length 10331;
  Best Local Similarity   98.6%;  Pred. No. 5.4e-186;
  Matches  614;  Conservative    0;  Mismatches    9;  Indels    0;  Gaps    0;

Qy         892 CAGCTCTTCAAAGGGAAGGTGAAGGTGGTGAGCCTGAAGCCTGAGGTGGCCCAGGTCGAC 951
               | |  ||||     | ||||||||||||||||||||||||||||||||||||||||||||
Db        6127 CTGTGCTTCTTCCTGCAGGTGAAGGTGGTGAGCCTGAAGCCTGAGGTGGCCCAGGTCGAC 6186

Qy         952 CTGTACATCCTCGGCCAAGCCGACCACTTTATTGGCAACTGTGTCTCCTCCTTCACTGCC 1011
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        6187 CTGTACATCCTCGGCCAAGCCGACCACTTTATTGGCAACTGTGTCTCCTCCTTCACTGCC 6246

Qy        1012 TTTGTGAAGCGGGAGCGGGACCTCCAGGGGAGGCCGTCTTCTTTCTTCGGCATGGACAGG 1071
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        6247 TTTGTGAAGCGGGAGCGGGACCTCCAGGGGAGGCCGTCTTCTTTCTTCGGCATGGACAGG 6306

Qy        1072 CCCCCTAAGCTGCGGGACGAGTTCTGATTCTGGCCGGAGCACCAGACCCTCTGATCCTGG 1131
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        6307 CCCCCTAAGCTGCGGGACGAGTTCTGATTCTGGCCGGAGCACCAGACCCTCTGATCCTGG 6366

Qy        1132 AGGGACCAGAGTCTGAGCTGGTCCTTCCAGCCAGGCCTGGCAGCCAGAGGTGCTCCGGGA 1191
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        6367 AGGGACCAGAGTCTGAGCTGGTCCTTCCAGCCAGGCCTGGCAGCCAGAGGTGCTCCGGGA 6426

Qy        1192 TTGCAAACTCCTCTTCTCACCTGCCAAAGATGGAGAAGAGTGCCAGGGACCCCTCAAGGA 1251
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        6427 TTGCAAACTCCTCTTCTCACCTGCCAAAGATGGAGAAGAGTGCCAGGGACCCCTCAAGGA 6486

Qy        1252 GGGAGACGCTCCATATCCCAGGGCATAGGACTTGCAGGTTCCTAGGAGCAGGAGCATCTC 1311
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        6487 GGGAGACGCTCCATATCCCAGGGCATAGGACTTGCAGGTTCCTAGGAGCAGGAGCATCTC 6546

Qy        1312 CCATCGCACGTGCTTTCTGCTCTTCTGGGAATTTCTCACACTGGCAAAGCAGTCCAGCCT 1371
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        6547 CCATCGCACGTGCTTTCTGCTCTTCTGGGAATTTCTCACACTGGCAAAGCAGTCCAGCCT 6606

Qy        1372 CCGTCTTCTGGTCCACTCTGCTCTGAGCAGCCTGGGATGCTGAACTCTTCAGAGAGATTT 1431
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        6607 CCGTCTTCTGGTCCACTCTGCTCTGAGCAGCCTGGGATGCTGAACTCTTCAGAGAGATTT 6666

Qy        1432 TTTTATAGAGAGATTTCTATAATTTTGATACAAGGTCATGACTATCCTAGAACTCTCTGT 1491
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        6667 TTTTATAGAGAGATTTCTATAATTTTGATACAAGGTCATGACTATCCTAGAACTCTCTGT 6726
```

```
Qy       1492 GGTTTTTGAAAATCATTGAATTC 1514
              |||||||||||||||||||||||
Db       6727 GGTTTTTGAAAATCATTGAATTC 6749


RESULT 9
US-10-956-157-9986
; Sequence 9986, Application US/10956157
; Publication No. US20050118625A1
; GENERAL INFORMATION:
;  APPLICANT: Wyeth
;  APPLICANT:  Mounts, William
;  TITLE OF INVENTION: NUCLEIC ACID ARRAYS FOR DETECTING GENE EXPRESSION ASSOCIATED WITH
;  TITLE OF INVENTION:  HUMAN OSTEOARTHRITIS AND HUMAN PROTEASES
;  FILE REFERENCE: 031896-043000 (AM 101081)
;  CURRENT APPLICATION NUMBER: US/10/956,157
;  CURRENT FILING DATE:  2004-10-04
;  NUMBER OF SEQ ID NOS: 319805
;  SOFTWARE: PatentIn version 3.2
; SEQ ID NO 9986
;   LENGTH: 1400
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-10-956-157-9986

  Query Match             31.1%;  Score 471.4;  DB 22;  Length 1400;
  Best Local Similarity   99.8%;  Pred. No. 8.7e-142;
  Matches  472;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 ATGCCCGCGGGCTCCTGGGACCCGGCCGGTTACCTGCTCTACTGCCCCTGCATGGGGCGC 60
              ||||| |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         13 ATGCCTGCGGGCTCCTGGGACCCGGCCGGTTACCTGCTCTACTGCCCCTGCATGGGGCGC 72

Qy         61 TTTGGGAACCAGGCCGATCACTTCTTGGGCTCTCTGGCATTTGCAAAGCTGCTAAACCGT 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         73 TTTGGGAACCAGGCCGATCACTTCTTGGGCTCTCTGGCATTTGCAAAGCTGCTAAACCGT 132

Qy        121 ACCTTGGCTGTCCCTCCTTGGATTGAGTACCAGCATCACAAGCCTCCTTTCACCAACCTC 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        133 ACCTTGGCTGTCCCTCCTTGGATTGAGTACCAGCATCACAAGCCTCCTTTCACCAACCTC 192

Qy        181 CATGTGTCCTACCAGAAGTACTTCAAGCTGGAGCCCCTCCAGGCTTACCATCGGGTCATC 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        193 CATGTGTCCTACCAGAAGTACTTCAAGCTGGAGCCCCTCCAGGCTTACCATCGGGTCATC 252

Qy        241 AGCTTGGAGGATTTCATGGAGAAGCTGGCACCCACCCACTGGCCCCCTGAGAAGCGGGTG 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        253 AGCTTGGAGGATTTCATGGAGAAGCTGGCACCCACCCACTGGCCCCCTGAGAAGCGGGTG 312

Qy        301 GCATACTGCTTTGAGGTGGCAGCCCAGCGAAGCCCAGATAAGAAGACGTGCCCCATGAAG 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        313 GCATACTGCTTTGAGGTGGCAGCCCAGCGAAGCCCAGATAAGAAGACGTGCCCCATGAAG 372

Qy        361 GAAGGAAACCCCTTTGGCCCATTCTGGGATCAGTTTCATGTGAGTTTCAACAAGTCGGAG 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        373 GAAGGAAACCCCTTTGGCCCATTCTGGGATCAGTTTCATGTGAGTTTCAACAAGTCGGAG 432

Qy        421 CTTTTTACAGGCATTTCCTTCAGTGCTTCCTACAGAGAACAATGGAGCCAGAG 473
              ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        433 CTTTTTACAGGCATTTCCTTCAGTGCTTCCTACAGAGAACAATGGAGCCAGAG 485


RESULT 10
US-10-956-157-4751
; Sequence 4751, Application US/10956157
; Publication No. US20050118625A1
; GENERAL INFORMATION:
;  APPLICANT: Wyeth
;  APPLICANT:  Mounts, William
;  TITLE OF INVENTION: NUCLEIC ACID ARRAYS FOR DETECTING GENE EXPRESSION ASSOCIATED WITH
;  TITLE OF INVENTION:  HUMAN OSTEOARTHRITIS AND HUMAN PROTEASES
;  FILE REFERENCE: 031896-043000 (AM 101081)
;  CURRENT APPLICATION NUMBER: US/10/956,157
;  CURRENT FILING DATE:  2004-10-04
;  NUMBER OF SEQ ID NOS: 319805
;  SOFTWARE: PatentIn version 3.2
; SEQ ID NO 4751
;   LENGTH: 1506
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-10-956-157-4751

  Query Match             31.1%;  Score 471.4;  DB 22;  Length 1506;
  Best Local Similarity   99.8%;  Pred. No. 9e-142;
  Matches  472;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 ATGCCCGCGGGCTCCTGGGACCCGGCCGGTTACCTGCTCTACTGCCCCTGCATGGGGCGC 60
              ||||| |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        119 ATGCCTGCGGGCTCCTGGGACCCGGCCGGTTACCTGCTCTACTGCCCCTGCATGGGGCGC 178

Qy         61 TTTGGGAACCAGGCCGATCACTTCTTGGGCTCTCTGGCATTTGCAAAGCTGCTAAACCGT 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        179 TTTGGGAACCAGGCCGATCACTTCTTGGGCTCTCTGGCATTTGCAAAGCTGCTAAACCGT 238

Qy        121 ACCTTGGCTGTCCCTCCTTGGATTGAGTACCAGCATCACAAGCCTCCTTTCACCAACCTC 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        239 ACCTTGGCTGTCCCTCCTTGGATTGAGTACCAGCATCACAAGCCTCCTTTCACCAACCTC 298

Qy        181 CATGTGTCCTACCAGAAGTACTTCAAGCTGGAGCCCCTCCAGGCTTACCATCGGGTCATC 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        299 CATGTGTCCTACCAGAAGTACTTCAAGCTGGAGCCCCTCCAGGCTTACCATCGGGTCATC 358

Qy        241 AGCTTGGAGGATTTCATGGAGAAGCTGGCACCCACCCACTGGCCCCCTGAGAAGCGGGTG 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        359 AGCTTGGAGGATTTCATGGAGAAGCTGGCACCCACCCACTGGCCCCCTGAGAAGCGGGTG 418

Qy        301 GCATACTGCTTTGAGGTGGCAGCCCAGCGAAGCCCAGATAAGAAGACGTGCCCCATGAAG 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        419 GCATACTGCTTTGAGGTGGCAGCCCAGCGAAGCCCAGATAAGAAGACGTGCCCCATGAAG 478

Qy        361 GAAGGAAACCCCTTTGGCCCATTCTGGGATCAGTTTCATGTGAGTTTCAACAAGTCGGAG 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        479 GAAGGAAACCCCTTTGGCCCATTCTGGGATCAGTTTCATGTGAGTTTCAACAAGTCGGAG 538

Qy        421 CTTTTTACAGGCATTTCCTTCAGTGCTTCCTACAGAGAACAATGGAGCCAGAG 473
              ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        539 CTTTTTACAGGCATTTCCTTCAGTGCTTCCTACAGAGAACAATGGAGCCAGAG 591


RESULT 11
US-09-918-995-22744
; Sequence 22744, Application US/09918995
; Publication No. US20030073623A1
; GENERAL INFORMATION:
;  APPLICANT: Hyseq, Inc.
;  TITLE OF INVENTION: NOVEL NUCLEIC ACID SEQUENCES OBTAINED
;  TITLE OF INVENTION:  FROM VARIOUS cDNA LIBRARIES
;  FILE REFERENCE: 20411-756
;  CURRENT APPLICATION NUMBER: US/09/918,995
;  CURRENT FILING DATE:  2001-07-30
;  PRIOR APPLICATION NUMBER: US/09/235,076
;  PRIOR FILING DATE: 1999-01-20
;  NUMBER OF SEQ ID NOS: 38054
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 22744
;   LENGTH: 477
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(477)
;   OTHER INFORMATION: n = A,T,C or G
US-09-918-995-22744

  Query Match             29.2%;  Score 442.8;  DB 10;  Length 477;
```

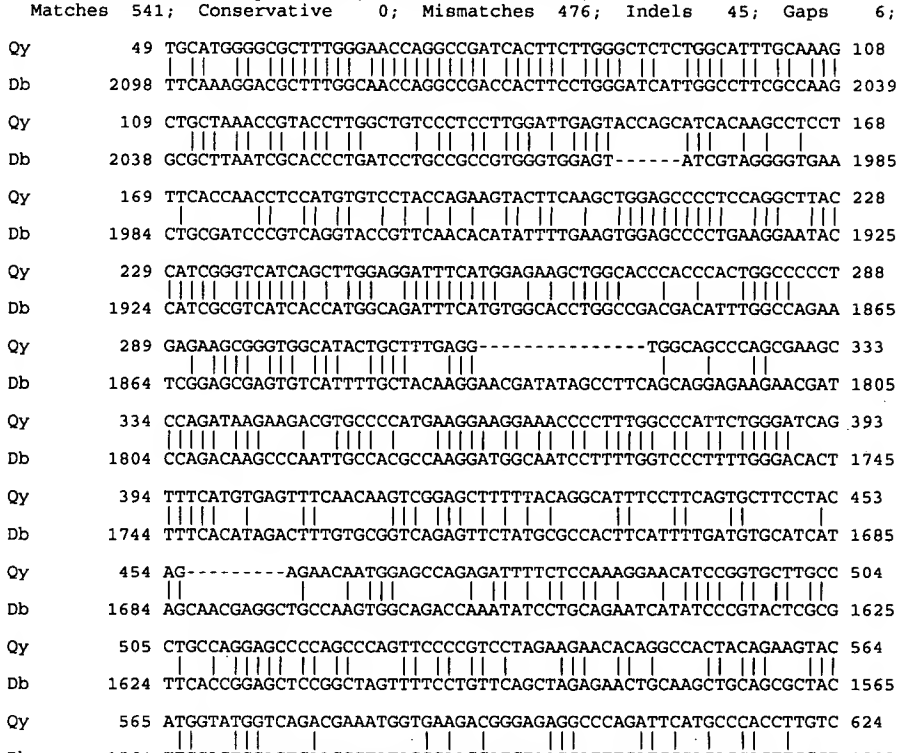
```
  Best Local Similarity   98.1%;  Pred. No. 1e-132;
  Matches  458;  Conservative    0;  Mismatches    8;  Indels    1;  Gaps    1;

Qy        855 TGCTACTGATTCCGAGAGTTATGTGCCTGAGCTCCAACAGCTCTTCAAAGGGAAGGTGAA 914
              || | ||||    | |||| ||||||||||||||| |||| ||||||||||||||||||
Db         11 TGATGCTGAGCNCTAGAGATATGTGCCTGAGCTCC-ACAGTTCTTCAAAGGGAAGGTGAA 69

Qy        915 GGTGGTGAGCCTGAAGCCTGAGGTGGCCCAGGTCGACCTGTACATCCTCGGCCAAGCCGA 974
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         70 GGTGGTGAGCCTGAAGCCTGAGGTGGCCCAGGTCGACCTGTACATCCTCGGCCAAGCCGA 129

Qy        975 CCACTTTATTGGCAACTGTGTCTCCTCCTTCACTGCCTTTGTGAAGCGGGAGCGGGACCT 1034
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        130 CCACTTTATTGGCAACTGTGTCTCCTCCTTCACTGCCTTTGTGAAGCGGGAGCGGGACCT 189

Qy       1035 CCAGGGGAGGCCGTCTTCTTTCTTCGGCATGGACAGGCCCCCTAAGCTGCGGGACGAGTT 1094
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        190 CCAGGGGAGGCCGTCTTCTTTCTTCGGCATGGACAGGCCCCCTAAGCTGCGGGACGAGTT 249

Qy       1095 CTGATTCTGGCCGGAGCACCAGACCCTCTGATCCTGGAGGGACCAGAGTCTGAGCTGGTC 1154
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        250 CTGATTCTGGCCGGAGCACCAGACCCTCTGATCCTGGAGGGACCAGAGTCTGAGCTGGTC 309

Qy       1155 CTTCCAGCCAGGCCTGGCAGCCAGAGGTGCTCCGGGATTGCAAACTCCTCTTCTCACCTG 1214
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        310 CTTCCAGCCAGGCCTGGCAGCCAGAGGTGCTCCGGGATTGCAAACTCCTCTTCTCACCTG 369

Qy       1215 CCAAAGATGGAGAAGAGTGCCAGGGACCCCTCAAGGAGGGAGACGCTCCATATCCCAGGG 1274
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        370 CCAAAGATGGAGAAGAGTGCCAGGGACCCCTCAAGGAGGGAGACGCTCCATATCCCAGGG 429

Qy       1275 CATAGGACTTGCAGGTTCCTAGGAGCAGGAGCATCTCCCATCGCACG 1321
              |||||||||||||||||||||||||||||||||||||||||||||||
Db        430 CATAGGACTTGCAGGTTCCTAGGAGCAGGAGCATCTCCCATCGCACG 476


RESULT 12
US-09-918-995-19437
; Sequence 19437, Application US/09918995
; Publication No. US20030073623A1
; GENERAL INFORMATION:
;  APPLICANT: Hyseq, Inc.
;  TITLE OF INVENTION: NOVEL NUCLEIC ACID SEQUENCES OBTAINED
;  TITLE OF INVENTION:  FROM VARIOUS cDNA LIBRARIES
;  FILE REFERENCE: 20411-756
;  CURRENT APPLICATION NUMBER: US/09/918,995
;  CURRENT FILING DATE:  2001-07-30
;  PRIOR APPLICATION NUMBER: US/09/235,076
;  PRIOR FILING DATE: 1999-01-20
;  NUMBER OF SEQ ID NOS: 38054
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 19437
;   LENGTH: 479
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-09-918-995-19437

  Query Match            29.1%;  Score 441;  DB 10;  Length 479;
  Best Local Similarity   98.9%;  Pred. No. 4e-132;
  Matches  444;  Conservative    0;  Mismatches    5;  Indels    0;  Gaps    0;

Qy        868 GAGAGTTATGTGCCTGAGCTCCAACAGCTCTTCAAAGGGAAGGTGAAGGTGGTGAGCCTG 927
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         31 GAGAGTTATGTGCCTGAGCTCCAACAGCTCTTCAAAGGGAAGGTGAAGGTGGTGAGCCTG 90

Qy        928 AAGCCTGAGGTGGCCCAGGTCGACCTGTACATCCTCGGCCAAGCCGACCACTTTATTGGC 987
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         91 AAGCCTGAGGTGGCCCAGGTCGACCTGTACATCCTCGGCCAAGCCGACCACTTTATTGGC 150

Qy        988 AACTGTGTCTCCTCCTTCACTGCCTTTGTGAAGCGGGAGCGGGACCTCCAGGGGAGGCCG 1047
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        151 AACTGTGTCTCCTCCTTCACTGCCTTTGTGAAGCGGGAGCGGGACCTCCAGGGGAGGCCG 210

Qy       1048 TCTTCTTTCTTCGGCATGGACAGGCCCCCTAAGCTGCGGGACGAGTTCTGATTCTGGCCG 1107
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        211 TCTTCTTTCTTCGGCATGGACAGGCCCCCTAAGCTGCGGGACGAGTTCTGATTCTGGCCG 270

Qy       1108 GAGCACCAGACCCTCTGATCCTGGAGGGACCAGAGTCTGAGCTGGTCCTTCCAGCCAGGC 1167
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        271 GAGCACCAGACCCTCTGATCCTGGAGGGACCAGAGTCTGAGCTGGTCCTTCCAGCCAGGC 330

Qy       1168 CTGGCAGCCAGAGGTGCTCCGGGATTGCAAACTCCTCTTCTCACCTGCCAAAGATGGAGA 1227
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        331 CTGGCAGCCAGAGGTGCTCCGGGATTGCAAACTCCTCTTCTCACCTGCCAAAGATGGAGA 390

Qy       1228 AGAGTGCCAGGGACCCCTCAAGGAGGGAGACGCTCCATATCCCAGGGCATAGGACTTGCA 1287
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        391 AGAGTGCCAGGGACCCCTCAAGGAGGGAGACGCTCCATATCCCAGGGCATAGGAACTTGA 450

Qy       1288 GGTTCCTAGGAGCAGGAGCATCTCCCATC 1316
              |||||||| ||||||||||||||||||||
Db        451 GGTTCCTAAGAGCAGGAGCATCTCCCATC 479


RESULT 13
US-11-097-143-18161
; Sequence 18161, Application US/11097143
; Publication No. US20050208558A1
; GENERAL INFORMATION:
;  APPLICANT: Venter, J. Craig
;  APPLICANT:  et al.
;  TITLE OF INVENTION: DETECTION KIT, SUCH AS NUCLEIC ACID
;  TITLE OF INVENTION:  ARRAYS, FOR DETECTING EXPRESSION OF 10,000 OR MORE
;  TITLE OF INVENTION:  DROSOPHILA GENES.
;  FILE REFERENCE: CL000728
;  CURRENT APPLICATION NUMBER: US/11/097,143
;  CURRENT FILING DATE:  2005-04-04
;  PRIOR APPLICATION NUMBER: 60/157,832
;  PRIOR FILING DATE: 1999-10-05
;  PRIOR APPLICATION NUMBER: 60/160,191
;  PRIOR FILING DATE: 1999-10-19
;  PRIOR APPLICATION NUMBER: 60/161,932
;  PRIOR FILING DATE: 1999-10-28
;  PRIOR APPLICATION NUMBER: 60/164,769
;  PRIOR FILING DATE: 1999-11-12
;  PRIOR APPLICATION NUMBER: 60/173,383
;  PRIOR FILING DATE: 1999-12-28
;  PRIOR APPLICATION NUMBER: 60/175,693
;  PRIOR FILING DATE: 2000-01-12
;  PRIOR APPLICATION NUMBER: 60/184,831
;  PRIOR FILING DATE: 2000-02-24
;  PRIOR APPLICATION NUMBER: 60/191,637
;  PRIOR FILING DATE: 2000-03-23
;  NUMBER OF SEQ ID NOS: 43008
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 18161
;   LENGTH: 1209
;   TYPE: DNA
;   ORGANISM: DROSOPHILA
US-11-097-143-18161

  Query Match            11.0%;  Score 166.4;  DB 26;  Length 1209;
  Best Local Similarity   51.5%;  Pred. No. 1.4e-42;
  Matches  563;  Conservative    0;  Mismatches  486;  Indels   45;  Gaps    6;

Qy         17 GGGACCCGGCCGGTTACCTGCTCTACTGCCCCTGCATGGGGCGCTTTGGGAACCAGGCCG 76
              | || ||    || |||||   |||||| || || ||||| ||||||| ||||||||||
Db         80 GCGATCCCAATGGCTACCTCACCTACTGTCCGTGTATGGGACGCTTTGGCAACCAGGCCG 139

Qy         77 ATCACTTCTTGGGCTCTCTGGCATTTGCAAAGCTGCTAAACCGTACCTTGGCTGTCCCTC 136
              | |||||| |||| ||  |||| || || |||  ||| || || ||| ||    | || |
Db        140 ACCACTTCCTGGGATCATTGGCCTTCGCCAAGGCGCTTAATCGCACCCTGATCCTGCCGC 199

Qy        137 CTTGGATTGAGTACCAGCATCACAAGCCTCCTTTCACCAACCTCCATGTGTCCTACCAGA 196
              | ||| | ||||      |||  | |  |    |       ||  || ||  | | | | |
Db        200 CGTGGGTGGAGT------ATCGTAGGGGTGAACTGCGATCCCGTCAGGTACCGTTCAACA 253
```



RESULT 14
US-11-097-143-18160/c
; Sequence 18160, Application US/11097143
; Publication No. US20050208558A1
; GENERAL INFORMATION:
; APPLICANT: Venter, J. Craig
; APPLICANT: et al.

; TITLE OF INVENTION: DETECTION KIT, SUCH AS NUCLEIC ACID
; TITLE OF INVENTION: ARRAYS, FOR DETECTING EXPRESSION OF 10,000 OR MORE
; TITLE OF INVENTION: DROSOPHILA GENES.
; FILE REFERENCE: CL000728
; CURRENT APPLICATION NUMBER: US/11/097,143
; CURRENT FILING DATE: 2005-04-04
; PRIOR APPLICATION NUMBER: 60/157,832
; PRIOR FILING DATE: 1999-10-05
; PRIOR APPLICATION NUMBER: 60/160,191
; PRIOR FILING DATE: 1999-10-19
; PRIOR APPLICATION NUMBER: 60/161,932
; PRIOR FILING DATE: 1999-10-28
; PRIOR APPLICATION NUMBER: 60/164,769
; PRIOR FILING DATE: 1999-11-12
; PRIOR APPLICATION NUMBER: 60/173,383
; PRIOR FILING DATE: 1999-12-28
; PRIOR APPLICATION NUMBER: 60/175,693
; PRIOR FILING DATE: 2000-01-12
; PRIOR APPLICATION NUMBER: 60/184,831
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: 60/191,637
; PRIOR FILING DATE: 2000-03-23
; NUMBER OF SEQ ID NOS: 43008
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 18160
; LENGTH: 3264
; TYPE: DNA
; ORGANISM: DROSOPHILA
US-11-097-143-18160

Query Match 9.9%; Score 150.4; DB 26; Length 3264;
Best Local Similarity 50.9%; Pred. No. 3.6e-37;
Matches 541; Conservative 0; Mismatches 476; Indels 45; Gaps 6;



```
Qy        625 CGG---CCCTATGTGGGCATTCATCTGCGCATTGGCTCTGACTGGAAGAACGCCTGTGCC 681
              |||   ||| | |||||||||||||||||||  ||   || |||  ||  |||||||
Db       1504 CGGGGTGCCTTTTTGGGCATTCATCTGCGCAACGGTATCGATTGGGTGAGAGCCTGTGAG 1445

Qy        682 ATGCTGAAGGACGG---------GACTGCAGGCTCGCACTTCATGGCCTCTCCGCAGTGT 732
              | |||||  |         |  |||     |||| | | ||| || |     |
Db       1444 CACGTCAAGGATAGCCAGCATCTGTTTGCCTCGCCGCAGTGCTTGGGCTATAAAAATGAA 1385

Qy        733 GTGGGCTACAGCCGCAGCACAGCGGCCCCCCTCACGATGACTATGTGCCTGCCTGACCTG 792
              ||      |  |         |  | | | | |     |  | |  |   | ||
Db       1384 CGTGGTGCACTCTACCCGGAGCTCTGCATGCCCTCCAAGGAGGCGATCATCCGCCAGCTA 1325

Qy        793 AAGGAGATCCAGAGGGCTGTGAAGCTCTGGGTGAGGTCGCTGGATGCCCAGTCGGTCTAC 852
              |||   | |   | |   |||   |       | |       ||   | | || || | |
Db       1324 AAGAGAACCATTAAGAACGTGCGCCAAACTCAGCCGGACAACGAAATCAAATCAGTTTTC 1265

Qy        853 GTTGCTACTGATTCCGAGAGTTATGTGCCTGAGCTCCAACAGCTCTTCAAAGGGAAGGTG 912
              || ||  | || ||| |         |   ||| ||  |   |  | | |   | | ||
Db       1264 GTGGCGTCAGACTCCAATCACATGATTGGTGAACTAAACACGGCCCTTAGTCGCATGGGC 1205

Qy        913 AAGGTGGTGAGCCTGAAGCCTGAGGTGG---CCCAGGTCGACCTGTACATCCTCGGCCAA 969
              |     |||  |  | ||| ||||  |   |  |  | ||| ||  ||| ||||| ||
Db       1204 ATCAGTGTGCACAAGCTGCCGGAGGATGATCCTTACCTGGACTTGGCCATTCTCGGACAG 1145

Qy        970 GCCGACCACTTTATTGGCAACTGTGTCTCCTCCTTCACTGCCTTTGTGAAGCGGGAGCGG 1029
              |  |||||||||| |||||||| | ||||| | |  || || |  ||  |||| ||
Db       1144 TCGAACCACTTTATCGGCAACTGTATATCCTCTTACTCGGCATTCGAAAAAAGGGAACGA 1085

Qy       1030 GACCTCCAGGGGAGGCCGTCTTCTTTCTTCGGCATGGACAGG 1071
              ||  | || ||    || || |  ||||  ||  |   || |
Db       1084 GATGTGCACGGTTTTCCATCGTACTTCTGGGGATTCCCCAAG 1043
```

RESULT 15
US-09-783-590-11501
; Sequence 11501, Application US/09783590
; Patent No. US20020110850A1
; GENERAL INFORMATION:
; APPLICANT: Dillon, Patrick J.
; APPLICANT: Haseltine, William A.
; APPLICANT: Li, Haodong
; APPLICANT: Rosen, Craig A.
; APPLICANT: Ruben, Steven M.
; TITLE OF INVENTION: Human Genes, Sequences, and Expression Products 16.2
; FILE REFERENCE: PO-16.2C1
; CURRENT APPLICATION NUMBER: US/09/783,590
; CURRENT FILING DATE: 2000-02-15
; PRIOR APPLICATION NUMBER: 08/420,856
; PRIOR FILING DATE: 1995-04-12
; PRIOR APPLICATION NUMBER: 08/346,731
; PRIOR FILING DATE: 1994-11-21
; NUMBER OF SEQ ID NOS: 12485
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 11501
; LENGTH: 402
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (9)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (19)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (22)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (46)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (272)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (301)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (320)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (333)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (342)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (374)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (386)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (390)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (400)
; OTHER INFORMATION: n equals a,t,g, or c
US-09-783-590-11501

Query Match 7.6%; Score 114.6; DB 9; Length 402;
Best Local Similarity 83.9%; Pred. No. 5.7e-26;
Matches 177; Conservative 0; Mismatches 23; Indels 11; Gaps 4;

```
Qy        178 CTCCATGTGTCCTACCAGAAGTACTTCAAGCTGGAGCCCCTCCAGGCTTACCATCGGGTC 237
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        161 CTCCATGTGTCCTACCAGAAGTACTTCAAGCTGGAGCCCCTCCAGGCTTACCATCGGGTC 220

Qy        238 ATCAGCTT-GGAGGATTTCATGGAGAAGCTGGCACCCACCCACTGGCCCCCT---GAGAA 293
              |||||||| |||||||||||||||||||||||||||||||||||||||| ||||    ||
Db        221 ATCAGCTTGGGAGGATTTCATGGAGAAGCTGGCACCCACCCACTGGGCCCCNTGAGAAGC 280

Qy        294 GCGGGTGGCATACTGCTTTGA--GGTGGCAGCCCAGCGAAGCCCAGAT-----AAGAAGA 346
              | ||  |||||| ||||||     |  |||||||||||| ||||||||     ||| |
Db        281 GGGGTGGGCATAATGCTTTTNAGGTGGGCAGCCCAGCGANGCCCAGATTAAGNAAGGACG 340

Qy        347 CGTGCCCCATGAAGGAAGGAAACCCCTTTGG 377
                  |||||||||||||||||| |||||| |
Db        341 TNGACCCCATGAAGGAAGGAAAACCCTTTTG 371
```

Search completed: November 9, 2005, 15:51:13
Job time : 1357.3 secs

GenCore version 5.1.6

OM nucleic - nucleic search, using sw model

Run on: November 9, 2005, 00:12:44 ; Search time 5490.13 Seconds
(without alignments)
10496.888 Million cell updates/sec

Title: US-09-774-954-1
Perfect score: 1514
Sequence: 1 ATGCCCGCGGGCTCCTGGGA.....TTTTGAAAATCATTGAATTC 1514

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 34239544 seqs, 19032134700 residues

Total number of hits satisfying chosen parameters: 68479088

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : EST:*
1: gb_est1:*
2: gb_est2:*
3: gb_htc:*
4: gb_est3:*
5: gb_est4:*
6: gb_est5:*
7: gb_est6:*
8: gb_gss1:*
9: gb_gss2:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	1443.8	95.4	1555	3	CR617950	full-leng
	2	892.8	59.0	1141	5	BX353274	BX353274 BX353274
	3	892	58.9	3611	3	AK081059	AK081059 Mus muscu
	4	888	58.7	2616	3	AK044629	AK044629 Mus muscu
c	5	859.6	56.8	889	5	BX345988	BX345988 BX345988
	6	852.4	56.3	914	5	BQ677046	BQ677046 AGENCOURT
	7	743.6	49.1	1091	5	BM909055	BM909055 AGENCOURT
	8	718.4	47.5	841	5	BX345989	BX345989 BX345989
	9	717.4	47.4	719	1	AL135434	AL135434 DKFZp762O
	10	712.2	47.0	889	4	BI524031	BI524031 603052202
	11	706.6	46.7	898	5	BX449172	BX449172 BX449172
	12	683.8	45.2	704	4	BM475212	BM475212 AGENCOURT
	13	673.2	44.5	942	7	CO774833	CO774833 ILLUMIGEN
	14	662.2	43.7	808	4	BG747760	BG747760 602705255
	15	660.2	43.6	801	5	BP144264	BP144264 BP144264
c	16	654.4	43.2	689	5	BM972332	BM972332 UI-CF-EC1
	17	639	42.2	988	5	BQ964179	BQ964179 AGENCOURT
	18	636.4	42.0	874	4	BI757540	BI757540 603029693
	19	630.8	41.7	947	4	BI410783	BI410783 602963556
	20	630.4	41.6	640	2	AW957639	AW957639 EST369709
	21	625.6	41.3	809	4	BG473023	BG473023 602514976
	22	624	41.2	915	5	BQ936634	BQ936634 AGENCOURT
	23	609.8	40.3	705	7	CN348925	CN348925 170005328
	24	607	40.1	608	4	BM711905	BM711905 UI-E-CL1-
	25	604.6	39.9	919	5	BU155142	BU155142 AGENCOURT
	26	599.2	39.6	753	4	BG424731	BG424731 602453575
	27	592	39.1	592	4	BM714731	BM714731 UI-E-EJ0-
	28	578.2	38.2	689	4	BI559635	BI559635 603252494
c	29	577.2	38.1	602	4	BM665708	BM665708 UI-E-CL1-
	30	558.4	36.9	561	4	BM798261	BM798261 K-EST0081
c	31	543.4	35.9	616	5	BX449171	BX449171 BX449171
	32	542.4	35.8	760	5	BU055535	BU055535 UI-M-FO0-
	33	541.8	35.8	545	7	CN348928	CN348928 170005313
	34	521.2	34.4	821	6	CD655308	CD655308 AGENCOURT
	35	511.2	33.8	763	7	CK366990	CK366990 AGENCOURT
	36	510	33.7	624	7	CN348922	CN348922 170004249
	37	503.8	33.3	800	2	BF036135	BF036135 601458122
	38	500.8	33.1	666	5	BP459416	BP459416 BP459416
	39	499.4	33.0	699	1	AU137536	AU137536 AU137536
	40	498	32.9	590	4	BG894786	BG894786 355525 MA
	41	494.2	32.6	736	5	BP457865	BP457865 BP457865
	42	491.8	32.5	769	1	AU117392	AU117392 AU117392
	43	485	32.0	644	7	CN459072	CN459072 UI-M-HN0-
	44	483.4	31.9	509	2	BE878197	BE878197 601487689
	45	473.4	31.3	4147	3	AK053648	AK053648 Mus muscu

ALIGNMENTS

RESULT 1
CR617950
LOCUS CR617950 1555 bp mRNA linear HTC 21-JUL-2004
DEFINITION full-length cDNA clone CS0DJ008YF12 of T cells (Jurkat cell line) Cot 10-normalized of Homo sapiens (human).
ACCESSION CR617950
VERSION CR617950.1 GI:50498757
KEYWORDS HTC; CNSLT_cDNA.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 1555)
AUTHORS Li,W.B., Gruber,C., Jessee,J. and Polayes,D.
TITLE Full-length cDNA libraries and normalization
JOURNAL Unpublished
REMARK Contact : Feng Liang Email : fliang@lifetech.com URL : http://fulllength.invitrogen.com/ InVitroGen Corporation 1600 Faraday Avenue
REFERENCE 2 (bases 1 to 1555)
AUTHORS Genoscope.
TITLE Direct Submission
JOURNAL Submitted (20-JUL-2004) Genoscope - Centre National de Sequencage : BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr - Web : www.genoscope.cns.fr)
COMMENT 1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime end enriched, double-strand cDNA was digested with Not I and cloned into the Not I and EcoR V sites of the pCMVSPORT 6 vector. Library was normalized. Library was constructed by Life Technologies, a division of Invitrogen.
FEATURES Location/Qualifiers
source 1. .1555
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="CS0DJ008YF12"
/tissue_type="T cells (Jurkat cell line) Cot 10-normalized"
/plasmid="pCMVSPORT_6"
ORIGIN

Query Match 95.4%; Score 1443.8; DB 3; Length 1555;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 1445; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 ATGCCCGCGGGCTCCTGGGACCCGGCCGGTTACCTGCTCTACTGCCCCTGCATGGGGCGC 60
||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

```
Db        109 ATGCCTGCGGGCTCCTGGGACCCGGCCGGTTACCTGCTCTACTGCCCCTGCATGGGGCGC 168
Qy         61 TTTGGGAACCAGGCCGATCACTTCTTGGGCTCTCTGGCATTTGCAAAGCTGCTAAACCGT 120
Db        169 TTTGGGAACCAGGCCGATCACTTCTTGGGCTCTCTGGCATTTGCAAAGCTGCTAAACCGT 228
Qy        121 ACCTTGGCTGTCCCTCCTTGGATTGAGTACCAGCATCACAAGCCTCCTTTCACCAACCTC 180
Db        229 ACCTTGGCTGTCCCTCCTTGGATTGAGTACCAGCATCACAAGCCTCCTTTCACCAACCTC 288
Qy        181 CATGTGTCCTACCAGAAGTACTTCAAGCTGGAGCCCCTCCAGGCTTACCATCGGGTCATC 240
Db        289 CATGTGTCCTACCAGAAGTACTTCAAGCTGGAGCCCCTCCAGGCTTACCATCGGGTCATC 348
Qy        241 AGCTTGGAGGATTTCATGGAGAAGCTGGCACCCACCCACTGGCCCCCTGAGAAGCGGGTG 300
Db        349 AGCTTGGAGGATTTCATGGAGAAGCTGGCACCCACCCACTGGCCCCCTGAGAAGCGGGTG 408
Qy        301 GCATACTGCTTTGAGGTGGCAGCCCAGCGAAGCCCAGATAAGAAGACGTGCCCCATGAAG 360
Db        409 GCATACTGCTTTGAGGTGGCAGCCCAGCGAAGCCCAGATAAGAAGACGTGCCCCATGAAG 468
Qy        361 GAAGGAAACCCCTTTGGCCCATTCTGGGATCAGTTTCATGTGAGTTTCAACAAGTCGGAG 420
Db        469 GAAGGAAACCCCTTTGGCCCATTCTGGGATCAGTTTCATGTGAGTTTCAACAAGTCGGAG 528
Qy        421 CTTTTTACAGGCATTTCCTTCAGTGCTTCCTACAGAGAACAATGGAGCCAGAGATTTTCT 480
Db        529 CTTTTTACAGGCATTTCCTTCAGTGCTTCCTACAGAGAACAATGGAGCCAGAGATTTTCT 588
Qy        481 CCAAAGGAACATCCGGTGCTTGCCCTGCCAGGAGCCCCAGCCCAGTTCCCCGTCCTAGAA 540
Db        589 CCAAAGGAACATCCGGTGCTTGCCCTGCCAGGAGCCCCAGCCCAGTTCCCCGTCCTAGAG 648
Qy        541 GAACACAGGCCACTACAGAAGTACATGGTATGGTCAGACGAAATGGTGAAGACGGGAGAG 600
Db        649 GAACACAGGCCACTACAGAAGTACATGGTATGGTCAGACGAAATGGTGAAGACGGGAGAG 708
Qy        601 GCCCAGATTCATGCCCACCTTGTCCGGCCCTATGTGGGCATTCATCTGCGCATTGGCTCT 660
Db        709 GCCCAGATTCATGCCCACCTTGTCCGGCCCTATGTGGGCATTCATCTGCGCATTGGCTCT 768
Qy        661 GACTGGAAGAACGCCTGTGCCATGCTGAAGGACGGGACTGCAGGCTCGCACTTCATGGCC 720
Db        769 GACTGGAAGAACGCCTGTGCCATGCTGAAGGACGGGACTGCAGGCTCGCACTTCATGGCC 828
Qy        721 TCTCCGCAGTGTGTGGGCTACAGCCGCAGCACAGCGGCCCCCCTCACGATGACTATGTGC 780
Db        829 TCTCCGCAGTGTGTGGGCTACAGCCGCAGCACAGCGGCCCCCCTCACGATGACTATGTGC 888
Qy        781 CTGCCTGACCTGAAGGAGATCCAGAGGGCTGTGAAGCTCTGGGTGAGGTCGCTGGATGCC 840
Db        889 CTGCCTGACCTGAAGGAGATCCAGAGGGCTGTGAAGCTCTGGGTGAGGTCGCTGGATGCC 948
Qy        841 CAGTCGGTCTACGTTGCTACTGATTCCGAGAGTTATGTGCCTGAGCTCCAACAGCTCTTC 900
Db        949 CAGTCGGTCTACGTTGCTACTGATTCCGAGAGTTATGTGCCTGAGCTCCAACAGCTCTTC 1008
Qy        901 AAAGGGAAGGTGAAGGTGGTGAGCCTGAAGCCTGAGGTGGCCCAGGTCGACCTGTACATC 960
Db       1009 AAAGGGAAGGTGAAGGTGGTGAGCCTGAAGCCTGAGGTGGCCCAGGTCGACCTGTACATC 1068
Qy        961 CTCGGCCAAGCCGACCACTTTATTGGCAACTGTGTCTCCTCCTTCACTGCCTTTGTGAAG 1020
Db       1069 CTCGGCCAAGCCGACCACTTTATTGGCAACTGTGTCTCCTCCTTCACTGCCTTTGTGAAG 1128
Qy       1021 CGGGAGCGGGACCTCCAGGGGAGGCCGTCTTCTTTCTTCGGCATGGACAGGCCCCCTAAG 1080
Db       1129 CGGGAGCGGGACCTCCAGGGGAGGCCGTCTTCTTTCTTCGGCATGGACAGGCCCCCTAAG 1188
Qy       1081 CTGCGGGACGAGTTCTGATTCTGGCCGGAGCACCAGACCCTCTGATCCTGGAGGGACCAG 1140
Db       1189 CTGCGGGACGAGTTCTGATTCTGGCCGGAGCACCAGACCCTCTGATCCTGGAGGGACCAG 1248
```

```
Qy       1141 AGTCTGAGCTGGTCCTTCCAGCCAGGCCTGGCAGCCAGAGGTGCTCCGGGATTGCAAACT 1200
Db       1249 AGTCTGAGCTGGTCCTTCCAGCCAGGCCTGGCAGCCAGAGGTGCTCCGGGATTGCAAACT 1308
Qy       1201 CCTCTTCTCACCTGCCAAAGATGGAGAAGAGTGCCAGGGACCCCTCAAGGAGGGAGACGC 1260
Db       1309 CCTCTTCTCACCTGCCAAAGATGGAGAAGAGTGCCAGGGACCCCTCAAGGAGGGAGACGC 1368
Qy       1261 TCCATATCCCAGGGCATAGGACTTGCAGGTTCCTAGGAGCAGGAGCATCTCCCATCGCAC 1320
Db       1369 TCCATATCCCAGGGCATAGGACTTGCAGGTTCCTAGGAGCAGGAGCATCTCCCATCGCAC 1428
Qy       1321 GTGCTTTCTGCTCTTCTGGGAATTTCTCACACTGGCAAAGCAGTCCAGCCTCCGTCTTCT 1380
Db       1429 GTGCTTTCTGCTCTTCTGGGAATTTCTCACACTGGCAAAGCAGTCCAGCCTCCGTCTTCT 1488
Qy       1381 GGTCCACTCTGCTCTGAGCAGCCTGGGATGCTGAACTCTTCAGAGAGATTTTTTTATAGA 1440
Db       1489 GGTCCACTCTGCTCTGAGCAGCCTGGGATGCTGAACTCTTCAGAGAGATTTTTTTATAGA 1548
Qy       1441 GAGATTT 1447
Db       1549 GAGATTT 1555
```

RESULT 2
BX353274

```
LOCUS       BX353274               1141 bp    mRNA    linear   EST 08-APR-2004
DEFINITION  BX353274 Homo sapiens NEUROBLASTOMA COT 25-NORMALIZED Homo sapiens
            cDNA clone CS0DC003YJ11 5-PRIME, mRNA sequence.
ACCESSION   BX353274
VERSION     BX353274.2  GI:46290116
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1141)
  AUTHORS   Li,W.B., Gruber,C., Jessee,J. and Polayes,D.
  TITLE     Full-length cDNA libraries and normalization
  JOURNAL   Unpublished (2001)
COMMENT     On May 5, 2003 this sequence version replaced gi:30371746.
            Contact: Genoscope
            Genoscope - Centre National de Sequencage
            2 rue Gaston Cremieux, CP 5706 - 91057 EVRY cedex - FRANCE
            Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr
            1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime
            end enriched, double-strand cDNA was digested with Not I and cloned
            into the Not I and EcoR V sites of the pCMVSPORT 6 vector. Library
            was normalized. Library was constructed by Life Technologies, a
            division of Invitrogen. This sequence belongs to sequence cluster
            9980.f
            For more information about this cluster, see
            http://www.genoscope.cns.fr/cdna?s=CS0DC003CE06QP1&c=9980.f.
FEATURES             Location/Qualifiers
     source          1. .1141
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="CS0DC003YJ11"
                     /tissue_type="NEUROBLASTOMA COT 25-NORMALIZED"
                     /clone_lib="Homo sapiens NEUROBLASTOMA COT 25-NORMALIZED"
                     /note="1st strand cDNA was primed with a NotI-oligo(dT)
                     primer. Five prime end enriched, double-strand cDNA was
                     digested with Not I and cloned into the Not I and EcoR V
                     sites of the pCMVSPORT 6 vector. Library was normalized."
ORIGIN
```

Query Match 59.0%; Score 892.8; DB 5; Length 1141;
Best Local Similarity 92.8%; Pred. No. 2.6e-236;
Matches 952; Conservative 32; Mismatches 35; Indels 7; Gaps 5;

```
Qy          1 ATGCCCGCGGGCTCCTGGGACCCGGCCGGTTACCTGCTCTACTGCCCCTGCATGGGGCGC 60
              ||||||||:|||||||||||||||||||||||||||||||||||||:|||||||||||||||
Db         86 ATGCCCGCRGGCTCCTGGGACCCGGCCGGTTACCTGCTCTACTGCMCCTGCATGGGGCGC 145

Qy         61 TTTGGGAACCAGGCCGATCACTTCTTGGGCTCTCTGGCATTTGCAAAGCTGCTAAACCGT 120
              ||||||||:|||||||||:|||||||||||||||||||||||||||||||:|||::||
Db        146 TTTGGGAAMCAGGCCGATMACTTCTTGGGCTCTCTGGCATTTGCAAAGCTGCWAAAMMGT 205

Qy        121 ACCTTGGCTGTCCCTCCTTGGATTGAGTACCAGCATCACAAGCCTCCTTTCACCAACCTC 180
              |||||||||| ||||||||||||||||||||||||||:||||:||||||||||||||||
Db        206 ACCTTGGCTGTACCTCCTTGGATTGAGTACCAGCATCAMAAGCMTCCTTTCACCAACCTC 265

Qy        181 CATGTGTCCTACCAGAAGTACTTCAAGCTGGAGCCCCTCCAGGCTTACCATCGGGTCATC 240
              ||||||||||||||||||||||:||:::|||||  |||:||||||||:||||||||||
Db        266 CATGTGTCCTACCAGAAGTACTTMAARMWGGAGCACCTMCAGGCTTACMATCGGGTCATC 325

Qy        241 AGCTTGGAGGATTTCATGGAGAAGCTGGCACCCACCCACTGGCCCCCTGAGAAGCGGGTG 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        326 AGCTTGGAGGATTTCATGGAGAAGCTGGCACCCACCCACTGGCCCCCTGAGAAGCGGGTG 385

Qy        301 GCATACTGCTTTGAGGTGGCAGCCCAGCGAAGCCCAGATAAGAAGACGTGCCCCATGAAG 360
              |||||||||||||||||||||:|||||||| |||||:||||||||||||||||||||
Db        386 GCATACTGCTTTGAGGTGGCAGCMCAGCGAAG-CCAGAWAAGAAGACGTGCCCCATGAAG 444

Qy        361 GAAGGAAACCCCTTTGGCCCATTCTGGGATCAGTTTCATGTGAGTTTCAACAAGTCGGAG 420
              |||||||||||||||||||||||||||||||||||||||||| ||||||||||||||||
Db        445 GAAGGAAACCCCTTTGGCCCATTCTGGGATCAGTTTCATGTGAGTATCAACAAGTCGGAG 504

Qy        421 CTTTTTACAGGCATTTCCTTCAGTGCTTCCTACAGAGAACAATGGAGCCAGAGATTTTCT 480
              ||||||||||||||||:||||||||||||||||||||||||||||||||||||||||||
Db        505 CTTTTTACAGGCATTTMCTTCAGTGCTTCCTACAGAGAACAATGGAGCCAGAGATTTTCT 564

Qy        481 CCAAAGGAACATCCGGTGCTTGCCCTGCCAGGAGCCCCAGCCCAGTTCCCCGTCCTAGAA 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        565 CCAAAGGAACATCCGGTGCTTGCCCTGCCAGGAGCCCCAGCCCAGTTCCCCGTCCTAGAG 624

Qy        541 GAACACAGGCCACTACAGAAGTACATGGTATGGTCAGACGAAATGGTGAAGACGGGAGAG 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||:
Db        625 GAACACAGGCCACTACAGAAGTACATGGTATGGTCAGACGAAATGGTGAAGACGGGAGAR 684

Qy        601 GCCCAGATTCATGCCCACCTTGTCCGGCCCTATGTGGGCATTCATCTGCGCATTGGCTCT 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        685 GCCCAGATTCATGCCCACCTTGTCCGGCCCTATGTGGGCATTCATCTGCGCATTGGCTCT 744

Qy        661 GACTGGAAGAACGCCTGTGCCATGCTGAAGGACGGGACTGCAGGCTCGCACTTCATGGCC 720
              ||||||||||||||||:|||||||||||||||||||||||||||||||||||||||||||
Db        745 GACTGGAAGAACGCCTGTDCCATGCTGAAGGACGGGACTGCAGGCTCGCACTTCATGGCC 804

Qy        721 TCTCCGCAGTGTGTGGGCTACAGCCGCAGCACAGCGGCCCCCCTCACGATGACTATGTGC 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        805 TCTCCGCAGTGTGTGGGCTACAGCCGCAGCACAGCGGCCCCCCTCACGATGACTATGTGC 864

Qy        781 CTGCCTGACCTGAAGGAGATCCAGAGGGCTGTGAAGCTCT-GGGTGAGGTCGCTGGATGC 839
              ||||||||||||||||||||||||||||||:||||||| ||||||||||||||||||||
Db        865 CTGCCTGACCTGAAGGAGATCCAGAGGGCTGTKAAGCTCTGGGGTGAGGTCGCTGGATGC 924

Qy        840 CCAGTCGGTCTACGTTGCTACTGATTCCGAGAGTTATGTGCCTGAGCTCCAACAGCTCTT 899
              |||||||||||||||||||||||||||||||||||| |||:||||||||||||||||||
Db        925 CCAGTCGGTCTACGTTGCTACTGATTCCGAGAGTTATGTTCCTKAGCTCCAACAGCTCTT 984

Qy        900 CAAAGGGAAGGTGAAGGTGGTGAGCCTGAAGCCTGAGGTGGCCCAGGTCGACCTGTACAT 959
              |||||||||||||||||||||||||| ||:||||:|||||:|:||   |     ||||||||
Db        985 CAAAGGGAAGGTGAAGGTGGTGAG-CTKAAGCYTGAGGKGSCC--AGGTCGACTGTAMAT 1041

Qy        960 CCTCGGCCAAGCCGACCACTTTATTGGCAACTGTGTCTCCTCCTTCACTGCCTTTGTGAA 1019
              |||||||  |  ||: |||||||||||||| : : |   |:   |  ||   |:   |
Db       1042 CCTCGGC--AAGCGRACACTTTATTGGCAATKTKTTTCTCYTTATTGCTTTGTKAAGGGG 1099

Qy       1020 GCGGGA 1025
                ||||
Db       1100 AGGGGA 1105
```

```
RESULT 3
AK081059
LOCUS       AK081059                3611 bp    mRNA    linear   HTC 03-APR-2004
DEFINITION  Mus musculus 10 days neonate cerebellum cDNA, RIKEN full-length
            enriched library, clone:B930076G19 product:protein
            O-fucosyltransferase 1, full insert sequence.
ACCESSION   AK081059
VERSION     AK081059.1  GI:26099647
KEYWORDS    HTC; CAP trapper.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1
  AUTHORS   Carninci,P. and Hayashizaki,Y.
  TITLE     High-efficiency full-length cDNA cloning
  JOURNAL   Meth. Enzymol. 303, 19-44 (1999)
  MEDLINE   99279253
   PUBMED   10349636
REFERENCE   2
  AUTHORS   Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K.,
            Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
  TITLE     Normalization and subtraction of cap-trapper-selected cDNAs to
            prepare full-length cDNA libraries for rapid discovery of new genes
  JOURNAL   Genome Res. 10 (10), 1617-1630 (2000)
  MEDLINE   20499374
   PUBMED   11042159
REFERENCE   3
  AUTHORS   Shibata,K., Itoh,M., Aizawa,K., Nagaoka,S., Sasaki,N., Carninci,P.,
            Konno,H., Akiyama,J., Nishi,K., Kitsunai,T., Tashiro,H., Itoh,M.,
            Sumi,N., Ishii,Y., Nakamura,S., Hazama,M., Nishine,T., Harada,A.,
            Yamamoto,R., Matsumoto,H., Sakaguchi,S., Ikegami,T., Kashiwagi,K.,
            Fujiwake,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E., Watahiki,M.,
            Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsuura,S., Kawai,J.,
            Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and Hayashizaki,Y.
  TITLE     RIKEN integrated sequence analysis (RISA) system--384-format
            sequencing pipeline with 384 multicapillary sequencer
  JOURNAL   Genome Res. 10 (11), 1757-1771 (2000)
  MEDLINE   20530913
   PUBMED   11076861
REFERENCE   4
  AUTHORS   The RIKEN Genome Exploration Research Group Phase II Team and the
            FANTOM Consortium.
  TITLE     Functional annotation of a full-length mouse cDNA collection
  JOURNAL   Nature 409, 685-690 (2001)
REFERENCE   5
  AUTHORS   The FANTOM Consortium and the RIKEN Genome Exploration Research
            Group Phase I & II Team.
  TITLE     Analysis of the mouse transcriptome based on functional annotation
            of 60,770 full-length cDNAs
  JOURNAL   Nature 420, 563-573 (2002)
REFERENCE   6  (bases 1 to 3611)
  AUTHORS   Adachi,J., Aizawa,K., Akimura,T., Arakawa,T., Bono,H., Carninci,P.,
            Fukuda,S., Furuno,M., Hanagaki,T., Hara,A., Hashizume,W.,
            Hayashida,K., Hayatsu,N., Hiramoto,K., Hiraoka,T., Hirozane,T.,
            Hori,F., Imotani,K., Ishii,Y., Itoh,M., Kagawa,I., Kasukawa,T.,
            Katoh,H., Kawai,J., Kojima,Y., Kondo,S., Konno,H., Kouda,M.,
            Koya,S., Kurihara,C., Matsuyama,T., Miyazaki,A., Murata,M.,
            Nakamura,M., Nishi,K., Nomura,K., Numazaki,R., Ohno,M., Ohsato,N.,
            Okazaki,Y., Saito,R., Saitoh,H., Sakai,C., Sakai,K., Sakazume,N.,
            Sano,H., Sasaki,D., Shibata,K., Shinagawa,A., Shiraki,T.,
            Sogabe,Y., Tagami,M., Tagawa,A., Takahashi,F., Takaku-Akahira,S.,
            Takeda,Y., Tanaka,T., Tomaru,A., Toya,T., Yasunishi,A.,
            Muramatsu,M. and Hayashizaki,Y.
  TITLE     Direct Submission
  JOURNAL   Submitted (16-APR-2002) Yoshihide Hayashizaki, The Institute of
            Physical and Chemical Research (RIKEN), Laboratory for Genome
            Exploration Research Group, RIKEN Genomic Sciences Center (GSC),
            RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama,
            Kanagawa 230-0045, Japan (E-mail:genome-res@gsc.riken.jp,
            URL:http://genome.gsc.riken.jp/, Tel:81-45-503-9222,
            Fax:81-45-503-9216)
COMMENT     cDNA library was prepared and sequenced in Mouse Genome
```

```
            Encyclopedia Project of Genome Exploration Research Group in Riken
            Genomic Sciences Center and Genome Science Laboratory in RIKEN.
            Division of Experimental Animal Research in Riken contributed to
            prepare mouse tissues.
            Please visit our web site for further details.
            URL:http://genome.gsc.riken.jp/
            URL:http://fantom.gsc.riken.jp/.
FEATURES             Location/Qualifiers
     source          1. .3611
                     /organism="Mus musculus"
                     /mol_type="mRNA"
                     /strain="C57BL/6J"
                     /db_xref="FANTOM_DB:B930076G19"
                     /db_xref="taxon:10090"
                     /clone="B930076G19"
                     /tissue_type="cerebellum"
                     /clone_lib="RIKEN full-length enriched mouse cDNA library"
                     /dev_stage="10 days neonate"
     misc_feature    11. .1194
                     /note="protein O-fucosyltransferase 1 (MGD|MGI:2153207,
                     GB|NM_080463, evidence: BLASTN, 99%, match=1290)
                     putative"
ORIGIN

  Query Match             58.9%;  Score 892;  DB 3;  Length 3611;
  Best Local Similarity   83.3%;  Pred. No. 6.3e-236;
  Matches 1051;  Conservative    0;  Mismatches  180;  Indels   31;  Gaps    2;

Qy          5 CCGCGGGCTCCTGGGACCCGGCCGGTTACCTGCTCTACTGCCCCTGCATGGGGCGCTTTG 64
              | ||||||||||||||||| ||||||||||||||||||||| ||||||||||||||||||
Db        101 CAGCGGGCTCCTGGGACCTGGCCGGTTACCTGCTCTACTGTCCCTGCATGGGGCGCTTTG 160

Qy         65 GGAACCAGGCCGATCACTTCTTGGGCTCTCTGGCATTTGCAAAGCTGCTAAACCGTACCT 124
              |||||||||| |||||||||||||||||| |||||||||| |||||||| ||||| ||||
Db        161 GGAACCAGGCTGATCACTTCTTGGGCTCCCTGGCATTTGCGAAGCTGCTGAACCGCACCT 220

Qy        125 TGGCTGTCCCTCCTTGGATTGAGTACCAGCATCACAAGCCTCCTTTCACCAACCTCCATG 184
              ||||||| ||||| |||||||| ||||| ||||||||||||||||||||||||||||||
Db        221 TGGCTGTACCTCCATGGATTGAATACCAACATCACAAGCCTCCTTTCACCAACCTCCATG 280

Qy        185 TGTCCTACCAGAAGTACTTCAAGCTGGAGCCCCTCCAGGCTTACCATCGGGTCATCAGCT 244
              |||||||||| |||||||||| |||||||| ||||| || |||||||||| |||||
Db        281 TGTCCTACCAAAAGTACTTCAAACTGGAGCCTCTCCAAGCCTACCATCGGGTTGTCAGCC 340

Qy        245 TGGAGGATTTCATGGAGAAGCTGGCACCCACCCACTGGCCCCCTGAGAAGCGGGTGGCAT 304
              ||||||| |||||||| || |||||||| ||||||||||||||||||||||||| |||||||
Db        341 TGGAGGACTTCATGGAAAATCTGGCACCCTCCCACTGGCCCCCTGAGAAGCGAGTGGCAT 400

Qy        305 ACTGCTTTGAGGTGGCAGCCCAGCGAAGCCCAGATAAGAAGACGTGCCCCATGAAGGAAG 364
              |||||||||||||||||||||||||||| || |||||||||| || ||||||||||||||
Db        401 ACTGCTTTGAGGTGGCAGCCCAGCGAAGTCCTGATAAGAAGACATGTCCCATGAAGGAAG 460

Qy        365 GAAACCCCTTTGGCCCATTCTGGGATCAGTTTCATGTGAGTTTCAACAAGTCGGAGCTTT 424
              |||| || ||||| |||||||||| ||||||||||||||||||||| ||||| || || |
Db        461 GAAATCCTTTTGGGCCATTCTGGGACCAGTTTCATGTGAGTTTCAATAAGTCAGAACTGT 520

Qy        425 TTACAGGCATTTCCTTCAGTGCTTCCTACAGAGAACAATGGAGCCAGAGATTTTCTCCAA 484
              | |||||||||||||||| || ||||||| |||||||||| |||||||||| || |||
Db        521 TCACAGGCATTTCCTTCAGCGCCTCCTACAAAGAACAATGGACCCAGAGATTTCCTGCAA 580

Qy        485 AGGAACATCCGGTGCTTGCCCTGCCAGGAGCCCCAGCCCAGTTCCCCGTCCTAGAAGAAC 544
              | || ||||| ||||| || ||||| || |||||||| |||||||| ||||| || ||||
Db        581 AAGAGCATCCTGTGCTCGCACTGCCTGGGGCCCCAGCACAGTTCCCTGTCCTGGAGGAAC 640

Qy        545 ACAGGCCACTACAGAAGTACATGGTATGGTCAGACGAAATGGTGAAGACGGGAGAGGCCC 604
              |||||   || ||||||||||||| |||||||| || |||||||| ||||||||||||||
Db        641 ACAGGGAGCTCCAGAAGTACATGGTGTGGTCAGATGAGATGGTGAGGACGGGAGAGGCCC 700

Qy        605 AGATTCATGCCCACCTTGTCCGGCCCTATGTGGGCATTCATCTGCGCATTGGCTCTGACT 664
              |||   |||||||| ||||||||||||||||||||||||||||||||||||||| ||||
Db        701 TGATCAGTGCCCACCTCGTCCGGCCCTATGTGGGCATTCATCTGCGCATTGGCTCCGACT 760

Qy        665 GGAAGAACGCCTGTGCCATGCTGAAGGACGGGACTGCAGGCTCGCACTTCATGGCCTCTC 724
              |||||||| |||||||||||||||||| || |||||||| || ||||||||||| || |
Db        761 GGAAGAATGCCTGTGCCATGCTGAAGGATGGAACTGCAGGGTCACACTTCATGGCTTCCC 820

Qy        725 CGCAGTGTGTGGGCTACAGCCGCAGCACAGCGGCCCCCCTCACGATGACTATGTGCCTGC 784
              | |||||||||||||| |||||||||||||||| |||| ||||| ||||| |||||||| |
Db        821 CTCAGTGTGTGGGCTATAGCCGCAGCACAGCGACCCCTCTCACCATGACCATGTGCCTCC 880

Qy        785 CTGACCTGAAGGAGATCCAGAGGGCTGTGAAGCTCTGGGTGAGGTCGCTGGATGCCCAGT 844
              |||||||||||||| |||||| |||||||||| ||| |||||||| | ||| |||||   |
Db        881 CTGACCTGAAGGAAATCCAGCGGGCTGTGACGCTTTGGGTGAGAGCACTGAATGCCAGAT 940

Qy        845 CGGTCTACGTTGCTACTGATTCCGAGAGTTATGTGCCTGAGCTCCAACAGCTCTTCAAAG 904
              |||||||| | || || || || ||||| || ||| | ||| |||| ||||||||||||||
Db        941 CGGTCTACATCGCCACAGACTCTGAGAGCTACGTGTCAGAGATCCAGCAGCTCTTCAAAG 1000

Qy        905 GGAAGGTGAAGGTGGTGAGCCTGAAGCCTGAGGTGGCCCAGGTCGACCTGTACATCCTCG 964
              |||||||| |||||||||||||||| || |||||||||||| ||||||||||||||||||
Db       1001 ACAAGGTGAGGGTGGTGAGCCTGAAACCCGAGGTGGCCCAGATCGACCTGTACATCCTCG 1060

Qy        965 GCCAAGCCGACCACTTTATTGGCAACTGTGTCTCCTCCTTCACTGCCTTTGTGAAGCGGG 1024
              |||| || |||||||| ||||| |||||||||||||| |||||||||| ||||||||||
Db       1061 GCCAGGCTGACCACTTCATTGGAAACTGTGTCTCCTCGTTCACTGCCTTCGTGAAGCGGG 1120

Qy       1025 AGCGGGACCTCCAGGGGAGGCCGTCTTCTTTCTTCGGCATGGACAGGCCCCCTAAGCTGC 1084
              |||||||||||||| |||||||| ||| || ||||| |||||||||| ||| |  |||| |
Db       1121 AGCGGGACCTCCATGGGAGGCAGTCGTCCTTCTTTGGCATGGACAGACCCTCCCAGCTTC 1180

Qy       1085 GGGACGAGTTCTGATTCTGGCCGGAGCACCAGACCCTCTGATCCTGGAGGGACCAGAGTC 1144
              |||| || || |||| |   | | ||| ||    | | |
Db       1181 GGGATGAATTTTGATCCCATCTGAAGCCCCCACCTCCC---------------------- 1218

Qy       1145 TGAGCTGGTCCTTCCAGCCAGGCCTGGCAGCCAGAGGTGCTCC-GGGATTGCAAACTCCT 1203
                    || || |||||||| |||| || |  | |||||| |||||| | |  | |||
Db       1219 --------ACCCTCAAGCCAGGCTTGGCTGCTAAGGATGCTCCTGGGATTACGCATCCCT 1270

Qy       1204 CTTCTCACCTGCCAAAGATGGAGAAGAGTGCCAGGGACCCCTCAAGGAGGGAGACGCTCC 1263
                ||||   |||||| || |||||||| ||| ||| ||||  |         | || |  |||
Db       1271 TTTCTTTTCTGCCAGAGGTGGAGAACAGTACCAAGGACGTCCTGGAAGAGAAGGCATTCC 1330

Qy       1264 AT 1265
                ||
Db       1331 AT 1332


RESULT 4
AK044629
LOCUS       AK044629               2616 bp    mRNA    linear   HTC 03-APR-2004
DEFINITION  Mus musculus adult retina cDNA, RIKEN full-length enriched library,
            clone:A930028F21 product:protein O-fucosyltransferase 1, full
            insert sequence.
ACCESSION   AK044629
VERSION     AK044629.1  GI:26336652
KEYWORDS    HTC; CAP trapper.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1
  AUTHORS   Carninci,P. and Hayashizaki,Y.
  TITLE     High-efficiency full-length cDNA cloning
  JOURNAL   Meth. Enzymol. 303, 19-44 (1999)
  MEDLINE   99279253
   PUBMED   10349636
REFERENCE   2
  AUTHORS   Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K.,
            Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
  TITLE     Normalization and subtraction of cap-trapper-selected cDNAs to
            prepare full-length cDNA libraries for rapid discovery of new genes
  JOURNAL   Genome Res. 10 (10), 1617-1630 (2000)
  MEDLINE   20499374
   PUBMED   11042159
REFERENCE   3
  AUTHORS   Shibata,K., Itoh,M., Aizawa,K., Nagaoka,S., Sasaki,N., Carninci,P.,
```

```
            Konno,H., Akiyama,J., Nishi,K., Kitsunai,T., Tashiro,H., Itoh,M.,
            Sumi,N., Ishii,Y., Nakamura,S., Hazama,M., Nishine,T., Harada,A.,
            Yamamoto,R., Matsumoto,H., Sakaguchi,S., Ikegami,T., Kashiwagi,K.,
            Fujiwake,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E., Watahiki,M.,
            Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsuura,S., Kawai,J.,
            Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and Hayashizaki,Y.
  TITLE     RIKEN integrated sequence analysis (RISA) system--384-format
            sequencing pipeline with 384 multicapillary sequencer
  JOURNAL   Genome Res. 10 (11), 1757-1771 (2000)
  MEDLINE   20530913
   PUBMED   11076861
REFERENCE   4
  AUTHORS   The RIKEN Genome Exploration Research Group Phase II Team and the
            FANTOM Consortium.
  TITLE     Functional annotation of a full-length mouse cDNA collection
  JOURNAL   Nature 409, 685-690 (2001)
REFERENCE   5
  AUTHORS   The FANTOM Consortium and the RIKEN Genome Exploration Research
            Group Phase I & II Team.
  TITLE     Analysis of the mouse transcriptome based on functional annotation
            of 60,770 full-length cDNAs
  JOURNAL   Nature 420, 563-573 (2002)
REFERENCE   6  (bases 1 to 2616)
  AUTHORS   Adachi,J., Aizawa,K., Akimura,T., Arakawa,T., Bono,H., Carninci,P.,
            Fukuda,S., Furuno,M., Hanagaki,T., Hara,A., Hashizume,W.,
            Hayashida,K., Hayatsu,N., Hiramoto,K., Hiraoka,T., Hirozane,T.,
            Hori,F., Imotani,K., Ishii,Y., Itoh,M., Kagawa,I., Kasukawa,T.,
            Katoh,H., Kawai,J., Kojima,Y., Kondo,S., Konno,H., Kouda,M.,
            Koya,S., Kurihara,C., Matsuyama,T., Miyazaki,A., Murata,M.,
            Nakamura,M., Nishi,K., Nomura,K., Numazaki,R., Ohno,M., Ohsato,N.,
            Okazaki,Y., Saito,R., Saitoh,H., Sakai,C., Sakai,K., Sakazume,N.,
            Sano,H., Sasaki,D., Shibata,K., Shinagawa,A., Shiraki,T.,
            Sogabe,Y., Tagami,M., Tagawa,A., Takahashi,F., Takaku-Akahira,S.,
            Takeda,Y., Tanaka,T., Tomaru,A., Toya,T., Yasunishi,A.,
            Muramatsu,M. and Hayashizaki,Y.
  TITLE     Direct Submission
  JOURNAL   Submitted (16-JUL-2001) Yoshihide Hayashizaki, The Institute of
            Physical and Chemical Research (RIKEN), Laboratory for Genome
            Exploration Research Group, RIKEN Genomic Sciences Center (GSC),
            RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama,
            Kanagawa 230-0045, Japan (E-mail:genome-res@gsc.riken.jp,
            URL:http://genome.gsc.riken.jp/, Tel:81-45-503-9222,
            Fax:81-45-503-9216)
COMMENT     cDNA library was prepared and sequenced in Mouse Genome
            Encyclopedia Project of Genome Exploration Research Group in Riken
            Genomic Sciences Center and Genome Science Laboratory in RIKEN.
            Division of Experimental Animal Research in Riken contributed to
            prepare mouse tissues.
            Retina RNA was provided by Dr. Stefano Gustincich (Department of
            Neurobiology, Harvard Medical School, 220 Longwood Ave., Boston, MA
            02115, USA) whose assistance is gratefully acknowledged. Please
            visit our web site for further details.
            URL:http://genome.gsc.riken.jp/
            URL:http://fantom.gsc.riken.jp/.
FEATURES             Location/Qualifiers
     source          1. .2616
                     /organism="Mus musculus"
                     /mol_type="mRNA"
                     /strain="C57BL/6J"
                     /db_xref="FANTOM_DB:A930028F21"
                     /db_xref="taxon:10090"
                     /clone="A930028F21"
                     /tissue_type="retina"
                     /clone_lib="RIKEN full-length enriched mouse cDNA library"
                     /dev_stage="adult"
     CDS             34. .1215
                     /note="unnamed protein product; protein
                     O-fucosyltransferase 1 (MGD|MGI:2153207, GB|NM_080463,
                     evidence: BLASTN, 99%, match=1290)
                     putative"
                     /codon_start=1
                     /protein_id="BAC32009.1"
                     /db_xref="GI:26336653"
                     /translation="MGAAAWAPPHLLLRASFLLLLLLLPLRGRSAGSWDLAGYLLYCP
                     CMGRFGNQADHFLGSLAFAKLLNRTLAVPPWIEYQHHKPPFTNLHVSYQKYFKLEPLQ
                     AYHRVVSLEDFMENLAPSHWPPEKRVAYCFEVAAQRSPDKKTCPMKEGNPFGPFWDQF
                     HVSFNKSELFTGISFSASYKEQWTQRFPAKEHPVLALPGAPAQFPVLEEHRELQKYMV
                     WSDEMVRTGEALISSHLVRPYVGIHLRIGSDWKNACAMLKDGTAGSHFMASPQCVGYS
                     RSTATPLTMTMCLPDLKEIQRAVTLWVRALNARSVYIATDSESYVSEIQQLFKDKVRV
                     VSLKPEVAQIDLYILGQADHFIGNCVSSFTAFVKRERDLHGRQSSFFGMDRPSQLRDE
                     F"
ORIGIN

  Query Match          58.7%;  Score 888;  DB 3;  Length 2616;
  Best Local Similarity  87.1%;  Pred. No. 7.4e-235;
  Matches  975;  Conservative    0;  Mismatches  145;  Indels    0;  Gaps    0;

Qy          3 GCCCGCGGGCTCCTGGGACCCGGCCGGTTACCTGCTCTACTGCCCCTGCATGGGGCGCTT 62
              | | |||||||||||||||| |||||||||||||||||||| ||||||||||||||||||
Db        120 GTCAGCGGGCTCCTGGGACCTGGCCGGTTACCTGCTCTACTGTCCCTGCATGGGGCGCTT 179

Qy         63 TGGGAACCAGGCCGATCACTTCTTGGGCTCTCTGGCATTTGCAAAGCTGCTAAACCGTAC 122
              |||||||||||| |||||||||||||||||||| |||||||||| |||||||| ||||| ||
Db        180 TGGGAACCAGGCTGATCACTTCTTGGGCTCCCTGGCATTTGCGAAGCTGCTGAACCGCAC 239

Qy        123 CTTGGCTGTCCCTCCTTGGATTGAGTACCAGCATCACAAGCCTCCTTTCACCAACCTCCA 182
              ||||||||| ||||| |||||||| ||||| ||||||||||||||||||||||||||||||
Db        240 CTTGGCTGTACCTCCATGGATTGAATACCAACATCACAAGCCTCCTTTCACCAACCTCCA 299

Qy        183 TGTGTCCTACCAGAAGTACTTCAAGCTGGAGCCCCTCCAGGCTTACCATCGGGTCATCAG 242
              |||||||||||| |||||||||| |||||||| ||||| || |||||||||||  ||||
Db        300 TGTGTCCTACCAAAAGTACTTCAAACTGGAGCCTCTCCAAGCCTACCATCGGGTTGTCAG 359

Qy        243 CTTGGAGGATTTCATGGAGAAGCTGGCACCCACCCACTGGCCCCCTGAGAAGCGGGTGGC 302
              | ||||||| |||||||| || |||||||| ||||||||||||||||||||||| |||||
Db        360 CCTGGAGGACTTCATGGAAAATCTGGCACCCTCCCACTGGCCCCCTGAGAAGCGAGTGGC 419

Qy        303 ATACTGCTTTGAGGTGGCAGCCCAGCGAAGCCCAGATAAGAAGACGTGCCCCATGAAGGA 362
              |||||||||||||||||||||||||||||| || |||||||||| || ||||||||||
Db        420 ATACTGCTTTGAGGTGGCAGCCCAGCGAAGTCCTGATAAGAAGACATGTCCCATGAAGGA 479

Qy        363 AGGAAACCCCTTTGGCCCATTCTGGGATCAGTTTCATGTGAGTTTCAACAAGTCGGAGCT 422
              |||||| || ||||| |||||||||| |||||||||||||||||||| ||||| || ||
Db        480 AGGAAATCCTTTTGGGCCATTCTGGGACCAGTTTCATGTGAGTTTCAATAAGTCAGAACT 539

Qy        423 TTTTACAGGCATTTCCTTCAGTGCTTCCTACAGAGAACAATGGAGCCAGAGATTTTCTCC 482
              || |||||||||||||||| || ||||||| |||||||||||| |||||||||||| || |
Db        540 GTTCACAGGCATTTCCTTCAGCGCCTCCTACAAAGAACAATGGACCCAGAGATTTCCTGC 599

Qy        483 AAAGGAACATCCGGTGCTTGCCCTGCCAGGAGCCCCAGCCCAGTTCCCCGTCCTAGAAGA 542
              ||| || ||||| ||||| || ||||| || |||||||| |||||||| ||||| || ||
Db        600 AAAAGAGCATCCTGTGCTCGCACTGCCTGGGGCCCCAGCACAGTTCCCTGTCCTGGAGGA 659

Qy        543 ACACAGGCCACTACAGAAGTACATGGTATGGTCAGACGAAATGGTGAAGACGGGAGAGGC 602
              |||||||   || |||||||||||||| |||||||| || ||||||| ||||||||||||
Db        660 ACACAGGGAGCTCCAGAAGTACATGGTGTGGTCAGATGAGATGGTGAGGACGGGAGAGGC 719

Qy        603 CCAGATTCATGCCCACCTTGTCCGGCCCTATGTGGGCATTCATCTGCGCATTGGCTCTGA 662
              ||| |||   | ||||||||||||||||||||||||||||||||||||||||||||| ||
Db        720 CCTGATCAGTTCCCACCTCGTCCGGCCCTATGTGGGCATTCATCTGCGCATTGGCTCCGA 779

Qy        663 CTGGAAGAACGCCTGTGCCATGCTGAAGGACGGGACTGCAGGCTCGCACTTCATGGCCTC 722
              ||||||||| |||||||||||||||||||| || |||||||| || |||||||||||  ||
Db        780 CTGGAAGAATGCCTGTGCCATGCTGAAGGATGGAACTGCAGGGTCACACTTCATGGCTTC 839

Qy        723 TCCGCAGTGTGTGGGCTACAGCCGCAGCACAGCGGCCCCCCTCACGATGACTATGTGCCT 782
              || |||||||||||||| |||||||||||||| |||| ||||| ||||| ||||||||
Db        840 CCCTCAGTGTGTGGGCTATAGCCGCAGCACAGCGACCCCTCTCACCATGACCATGTGCCT 899

Qy        783 GCCTGACCTGAAGGAGATCCAGAGGGCTGTGAAGCTCTGGGTGAGGTCGCTGGATGCCCA 842
              |||||||||||||||| |||||| ||| |||||||||  | ||| |||||
Db        900 CCCTGACCTGAAGGAAATCCAGCGGGCTGTGACGCTTTGGGTGAGAGCACTGAATGCCAG 959

Qy        843 GTCGGTCTACGTTGCTACTGATTCCGAGAGTTATGTGCCTGAGCTCCAACAGCTCTTCAA 902
              ||||||||| | || || || || ||||| || ||| | ||| |||| ||||||||||
Db        960 ATCGGTCTACATCGCCACAGACTCTGAGAGCTACGTGTCAGAGATCCAGCAGCTCTTCAA 1019
```

```
Qy        903 AGGGAAGGTGAAGGTGGTGAGCCTGAAGCCTGAGGTGGCCCAGGTCGACCTGTACATCCT 962
              ||  |||||||| ||||||||||||||| || ||||||||||||| ||||||||||||||||
Db       1020 AGACAAGGTGAGGGTGGTGAGCCTGAAACCCGAGGTGGCCCAGATCGACCTGTACATCCT 1079

Qy        963 CGGCCAAGCCGACCACTTTATTGGCAACTGTGTCTCCTCCTTCACTGCCTTTGTGAAGCG 1022
              |||||| || |||||||| ||||| |||||||||||||| |||||||||| ||||||||
Db       1080 CGGCCAGGCTGACCACTTCATTGGAAACTGTGTCTCCTCGTTCACTGCCTTCGTGAAGCG 1139

Qy       1023 GGAGCGGGACCTCCAGGGGAGGCCGTCTTCTTTCTTCGGCATGGACAGGCCCCCTAAGCT 1082
              |||||||||||||| ||||||| ||| || ||||| |||||||||| ||| |  ||||
Db       1140 GGAGCGGGACCTCCATGGGAGGCAGTCGTCCTTCTTTGGCATGGACAGACCCTCCCAGCT 1199

Qy       1083 GCGGGACGAGTTCTGATTCTGGCCGGAGCACCAGACCCTC 1122
                ||||| || || |||| |   | | ||| ||   | | |
Db       1200 TCGGGATGAATTTTGATCCCATCTGAAGCCCCCACCTCCC 1239
```

RESULT 5
BX345988/c

```
LOCUS       BX345988                889 bp    mRNA    linear   EST 08-APR-2004
DEFINITION  BX345988 Homo sapiens T CELLS (JURKAT CELL LINE) COT 10-NORMALIZED
            Homo sapiens cDNA clone CS0DJ008YF12 3-PRIME, mRNA sequence.
ACCESSION   BX345988
VERSION     BX345988.2  GI:46306019
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 889)
  AUTHORS   Li,W.B., Gruber,C., Jessee,J. and Polayes,D.
  TITLE     Full-length cDNA libraries and normalization
  JOURNAL   Unpublished (2001)
COMMENT     On May 5, 2003 this sequence version replaced gi:30375046.
            Contact: Genoscope
            Genoscope - Centre National de Sequencage
            2 rue Gaston Cremieux, CP 5706 - 91057 EVRY cedex - FRANCE
            Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr
            1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime
            end enriched, double-strand cDNA was digested with Not I and cloned
            into the Not I and EcoR V sites of the pCMVSPORT 6 vector. Library
            was normalized. Library was constructed by Life Technologies, a
            division of Invitrogen. This sequence belongs to sequence cluster
            9980.f
            For more information about this cluster, see
            http://www.genoscope.cns.fr/cdna?s=CS0AJ008DC06NP1&c=9980.f.
FEATURES             Location/Qualifiers
     source          1. .889
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="CS0DJ008YF12"
                     /cell_type="T CELLS (JURKAT CELL LINE) COT 10-NORMALIZED"
                     /cell_line="JURKAT"
                     /clone_lib="Homo sapiens T CELLS (JURKAT CELL LINE) COT
                     10-NORMALIZED"
                     /note="1st strand cDNA was primed with a NotI-oligo(dT)
                     primer. Five prime end enriched, double-strand cDNA was
                     digested with Not I and cloned into the Not I and EcoR V
                     sites of the pCMVSPORT 6 vector. Library was normalized."
ORIGIN
```

```
  Query Match            56.8%;  Score 859.6;  DB 5;  Length 889;
  Best Local Similarity  99.4%;  Pred. No. 4.1e-227;
  Matches  870;  Conservative    2;  Mismatches    2;  Indels    1;  Gaps    1;

Qy        573 GTCAGACGAAATGGTGAAGACGGGAGAGGCCCAGATTCATGCCCACCTTGTCCGGCCCTA 632
              |||||||||||| |||||||||||||||||||||||||||||||||||||:||||||
Db        874 GTCAGACGAAAT-GTGAAGACGGGAGAGGCCCAGATTCATGCCCACCTTGTCCGSCCCTA 816

Qy        633 TGTGGGCATTCATCTGCGCATTGGCTCTGACTGGAAGAACGCCTGTGCCATGCTGAAGGA 692
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

Db        815 TGTGGGCATTCATCTGCGCATTGGCTCTGACTGGAAGAACGCCTGTGCCATGCTGAAGGA 756

Qy        693 CGGGACTGCAGGCTCGCACTTCATGGCCTCTCCGCAGTGTGTGGGCTACAGCCGCAGCAC 752
              ||||||||||||:||||||||||||||||||||||||||||||||||||||||||||||
Db        755 CGGGACTGCAGGYTCGCACTTCATGGCCTCTCCGCAGTGTGTGGGCTACAGCCGCAGCAC 696

Qy        753 AGCGGCCCCCCCTCACGATGACTATGTGCCTGCCTGACCTGAAGGAGATCCAGAGGGCTGT 812
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        695 AGCGGCCCCCCCTCACGATGACTATGTGCCTGCCTGACCTGAAGGAGATCCAGAGGGCTGT 636

Qy        813 GAAGCTCTGGGTGAGGTCGCTGGATGCCCAGTCGGTCTACGTTGCTACTGATTCCGAGAG 872
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        635 GAAGCTCTGGGTGAGGTCGCTGGATGCCCAGTCGGTCTACGTTGCTACTGATTCCGAGAG 576

Qy        873 TTATGTGCCTGAGCTCCAACAGCTCTTCAAAGGGAAGGTGAAGGTGGTGAGCCTGAAGCC 932
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        575 TTATGTGCCTGAGCTCCAACAGCTCTTCAAAGGGAAGGTGAAGGTGGTGAGCCTGAAGCC 516

Qy        933 TGAGGTGGCCCAGGTCGACCTGTACATCCTCGGCCAAGCCGACCACTTTATTGGCAACTG 992
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        515 TGAGGTGGCCCAGGTCGACCTGTACATCCTCGGCCAAGCCGACCACTTTATTGGCAACTG 456

Qy        993 TGTCTCCTCCTTCACTGCCTTTGTGAAGCGGGAGCGGGACCTCCAGGGGAGGCCGTCTTC 1052
              |||||||||||| ||||||||||||||||||||||||||||||||||||||||||||||
Db        455 TGTCTCCTCCTTNACTGCCTTTGTGAAGCGGGAGCGGGACCTCCAGGGGAGGCCGTCTTC 396

Qy       1053 TTTCTTCGGCATGGACAGGCCCCCTAAGCTGCGGGACGAGTTCTGATTCTGGCCGGAGCA 1112
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        395 TTTCTTCGGCATGGACAGGCCCCCTAAGCTGCGGGACGAGTTCTGATTCTGGCCGGAGCA 336

Qy       1113 CCAGACCCTCTGATCCTGGAGGGACCAGAGTCTGAGCTGGTCCTTCCAGCCAGGCCTGGC 1172
              |||||||||||||||||||||||||||||||||||||||||| ||||||||||||||||||
Db        335 CCAGACCCTCTGATCCTGGAGGGACCAGAGTCTGAGCTGGTCTTTCCAGCCAGGCCTGGC 276

Qy       1173 AGCCAGAGGTGCTCCGGGATTGCAAACTCCTCTTCTCACCTGCCAAAGATGGAGAAGAGT 1232
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        275 AGCCAGAGGTGCTCCGGGATTGCAAACTCCTCTTCTCACCTGCCAAAGATGGAGAAGAGT 216

Qy       1233 GCCAGGGACCCCTCAAGGAGGGAGACGCTCCATATCCCAGGGCATAGGACTTGCAGGTTC 1292
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        215 GCCAGGGACCCCTCAAGGAGGGAGACGCTCCATATCCCAGGGCATAGGACTTGCAGGTTC 156

Qy       1293 CTAGGAGCAGGAGCATCTCCCATCGCACGTGCTTTCTGCTCTTCTGGGAATTTCTCACAC 1352
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        155 CTAGGAGCAGGAGCATCTCCCATCGCACGTGCTTTCTGCTCTTCTGGGAATTTCTCACAC 96

Qy       1353 TGGCAAAGCAGTCCAGCCTCCGTCTTCTGGTCCACTCTGCTCTGAGCAGCCTGGGATGCT 1412
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          95 TGGCAAAGCAGTCCAGCCTCCGTCTTCTGGTCCACTCTGCTCTGAGCAGCCTGGGATGCT 36

Qy       1413 GAACTCTTCAGAGAGATTTTTTTATAGAGAGATTT 1447
              |||||||||||||||||||||||||||||||||||
Db          35 GAACTCTTCAGAGAGATTTTTTTATAGAGAGATTT 1
```

RESULT 6
BQ677046

```
LOCUS       BQ677046                914 bp    mRNA    linear   EST 15-JUL-2002
DEFINITION  AGENCOURT_8343300 NIH_MGC_112 Homo sapiens cDNA clone IMAGE:6273476
            5', mRNA sequence.
ACCESSION   BQ677046
VERSION     BQ677046.1  GI:21789725
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 914)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
```

```
            Tissue Procurement: DCTD/DTP
             cDNA Library Preparation: Rubin Laboratory
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Agencourt Bioscience Corporation
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLCM2453  row: h  column: 21
            High quality sequence stop: 662.
FEATURES             Location/Qualifiers
     source          1. .914
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:6273476"
                     /tissue_type="melanotic melanoma, cell line"
                     /lab_host="DH10B (phage-resistant)"
                     /clone_lib="NIH_MGC_112"
                     /note="Organ: skin; Vector: pOTB7; Site_1: XhoI; Site_2:
                     EcoRI; cDNA made by oligo-dT priming. Directionally cloned
                     into EcoRI/XhoI sites using the following 5' adaptor:
                     GGCACGAG(G). Library constructed  by Ling Hong in the
                     laboratory of Gerald M. Rubin (University of  California,
                     Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and
                     Superscript II RT (Life Technologies). Note: this is a
                     NIH_MGC Library."
ORIGIN

  Query Match             56.3%;  Score 852.4;  DB 5;  Length 914;
  Best Local Similarity   99.7%;  Pred. No. 4.1e-225;
  Matches  864;  Conservative    0;  Mismatches    2;  Indels    1;  Gaps    1;

Qy        636 GGGCATTCATCTGCGCATTGGCTCTGACTGGAAGAACGCCTGTGCCATGCTGAAGGACGG 695
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 GGGCATTCATCTGCGCATTGGCTCTGACTGGAAGAACGCCTGTGCCATGCTGAAGGACGG 60

Qy        696 GACTGCAGGCTCGCACTTCATGGCCTCTCCGCAGTGTGTGGGCTACAGCCGCAGCACAGC 755
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GACTGCAGGCTCGCACTTCATGGCCTCTCCGCAGTGTGTGGGCTACAGCCGCAGCACAGC 120

Qy        756 GGCCCCCCTCACGATGACTATGTGCCTGCCTGACCTGAAGGAGATCCAGAGGGCTGTGAA 815
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 GGCCCCCCTCACGATGACTATGTGCCTGCCTGACCTGAAGGAGATCCAGAGGGCTGTGAA 180

Qy        816 GCTCTGGGTGAGGTCGCTGGATGCCCAGTCGGTCTACGTTGCTACTGATTCCGAGAGTTA 875
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 GCTCTGGGTGAGGTCGCTGGATGCCCAGTCGGTCTACGTTGCTACTGATTCCGAGAGTTA 240

Qy        876 TGTGCCTGAGCTCCAACAGCTCTTCAAAGGGAAGGTGAAGGTGGTGAGCCTGAAGCCTGA 935
              ||||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||
Db        241 TGTGCCTGAGCTCCAACAGTTCTTCAAAGGGAAGGTGAAGGTGGTGAGCCTGAAGCCTGA 300

Qy        936 GGTGGCCCAGGTCGACCTGTACATCCTCGGCCAAGCCGACCACTTTATTGGCAACTGTGT 995
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 GGTGGCCCAGGTCGACCTGTACATCCTCGGCCAAGCCGACCACTTTATTGGCAACTGTGT 360

Qy        996 CTCCTCCTTCACTGCCTTTGTGAAGCGGGAGCGGGACCTCCAGGGGAGGCCGTCTTCTTT 1055
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 CTCCTCCTTCACTGCCTTTGTGAAGCGGGAGCGGGACCTCCAGGGGAGGCCGTCTTCTTT 420

Qy       1056 CTTCGGCATGGACAGGCCCCCTAAGCTGCGGGACGAGTTCTGATTCTGGCCGGAGCACCA 1115
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 CTTCGGCATGGACAGGCCCCCTAAGCTGCGGGACGAGTTCTGATTCTGGCCGGAGCACCA 480

Qy       1116 GACCCTCTGATCCTGGAGGGACCAGAGTCTGAGCTGGTCCTTCCAGCCAGGCCTGGCAGC 1175
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 GACCCTCTGATCCTGGAGGGACCAGAGTCTGAGCTGGTCCTTCCAGCCAGGCCTGGCAGC 540

Qy       1176 CAGAGGTGCTCCGGGATTGCAAACTCCTCTTCTCACCTGCCAAAGATGGAGAAGAGTGCC 1235
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 CAGAGGTGCTCCGGGATTGCAAACTCCTCTTCTCACCTGCCAAAGATGGAGAAGAGTGCC 600

Qy       1236 AGGGACCCCTCAAGGAGGGAGACGCTCCATATCCCAGGGCATAGGACTTGCAGGTTCCTA 1295
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 AGGGACCCCTCAAGGAGGGAGACGCTCCATATCCCAGGGCATAGGACTTGCAGGTTCCTA 660

Qy       1296 GGAGCAGGAGCATCTCCCATCGCACGTGCTTTCTGCTCTTCTGGGAATTTCTCACACTGG 1355
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 GGAGCAGGAGCATCTCCCATCGCACGTGCTTTCTGCTCTTCTGGGAATTTCTCACACTGG 720

Qy       1356 CAAAGCAGTCCAGCCTCCGTCTTCTGGTCCACTCTGCTCTGAGCAGCCTGGGATGCTGAA 1415
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 CAAAGCAGTCCAGCCTCCGTCTTCTGGTCCACTCTGCTCTGAGCAGCCTGGGATGCTGAA 780

Qy       1416 CTCTTCAGAGAGATTTTTTTATAGAGAGATTTCTATAATTTTGATACAAGGTCATGACTA 1475
              |||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||
Db        781 CTCTTCAGAGAGATTTTTTTATAGAGAGA-TTCTATAATTTTGATACAAGGTCATGACTA 839

Qy       1476 TCCTAGAACTCTCTGTGGTTTTTGAAA 1502
              |||||||||||||||||| |||||||||
Db        840 TCCTAGAACTCTCTGTGGNTTTTGAAA 866


RESULT 7
BM909055
LOCUS       BM909055                1091 bp    mRNA    linear   EST 12-MAR-2002
DEFINITION  AGENCOURT_6611915 NIH_MGC_98 Homo sapiens cDNA clone IMAGE:5431669
            5', mRNA sequence.
ACCESSION   BM909055
VERSION     BM909055.1  GI:19359434
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1091)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: ATCC
             cDNA Library Preparation: Rubin Laboratory
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Agencourt Bioscience Corporation
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLCM1904  row: e  column: 14
            High quality sequence stop: 629.
FEATURES             Location/Qualifiers
     source          1. .1091
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:5431669"
                     /tissue_type="astrocytoma grade IV, cell line"
                     /lab_host="DH10B (phage-resistant)"
                     /clone_lib="NIH_MGC_98"
                     /note="Organ: brain; Vector: pOTB7; Site_1: XhoI; Site_2:
                     EcoRI; cDNA made by oligo-dT priming.  Directionally
                     cloned into EcoRI/XhoI sites using the following 5'
                     adaptor: GGCACGAG(G).  Library constructed  by Ling Hong
                     in the laboratory of Gerald M. Rubin (University of
                     California, Berkeley) using ZAP-cDNA synthesis kit
                     (Stratagene) and Superscript II RT (Life Technologies).
                     Note: this is a NIH_MGC Library."
ORIGIN

  Query Match             49.1%;  Score 743.6;  DB 5;  Length 1091;
  Best Local Similarity   93.0%;  Pred. No. 7.4e-195;
  Matches  824;  Conservative    0;  Mismatches   54;  Indels    8;  Gaps    4;

Qy        178 CTCCATGTGTCCTACCAGAAGTACTTCAAGCTGGAGCCCCTCCAGGCTTACCATCGGGTC 237
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        132 CTCCATGTGTCCTACCAGAAGTACTTCAAGCTGGAGCCCCTCCAGGCTTACCATCGGGTC 191
```

```
Qy        238 ATCAGCTTGGAGGATTTCATGGAGAAGCTGGCACCCACCCACTGGCCCCCTGAGAAGCGG 297
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        192 ATCAGCTTGGAGGATTTCATGGAGAAGCTGGCACCCACCCACTGGCCCCCTGAGAAGCGG 251

Qy        298 GTGGCATACTGCTTTGAGGTGGCAGCCCAGCGAAGCCCAGATAAGAAGACGTGCCCCATG 357
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        252 GTGGCATACTGCTTTGAGGTGGCAGCCCAGCGAAGCCCAGATAAGAAGACGTGCCCCATG 311

Qy        358 AAGGAAGGAAACCCCTTTGGCCCATTCTGGGATCAGTTTCATGTGAGTTTCAACAAGTCG 417
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        312 AAGGAAGGAAACCCCTTTGGCCCATTCTGGGATCAGTTTCATGTGAGTTTCAACAAGTCG 371

Qy        418 GAGCTTTTTACAGGCATTTCCTTCAGTGCTTCCTACAGAGAACAATGGAGCCAGAGATTT 477
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        372 GAGCTTTTTACAGGCATTTCCTTCAGTGCTTCCTACAGAGAACAATGGAGCCAGAGATTT 431

Qy        478 TCTCCAAAGGAACATCCGGTGCTTGCCCTGCCAGGAGCCCCAGCCCAGTTCCCCGTCCTA 537
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        432 TCTCCAAAGGAACATCCGGTGCTTGCCCTGCCAGGAGCCCCAGCCCAGTTCCCCGTCCTA 491

Qy        538 GAAGAACACAGGCCACTACAGAAGTACATGGTATGGTCAGACGAAATGGTGAAGACGGGA 597
              || ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        492 GAGGAACACAGGCCACTACAGAAGTACATGGTATGGTCAGACGAAATGGTGAAGACGGGA 551

Qy        598 GAGGCCCAGATTCATGCCCACCTTGTCCGGCCCTATGTGGGCATTCATCTGCGCATTGGC 657
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        552 GAGGCCCAGATTCATGCCCACCTTGTCCGGCCCTATGTGGGCATTCATCTGCGCATTGGC 611

Qy        658 TCTGACTGGAAGAACGCCTGTGCCATGCTGAAGGACGGGACTGCAGGCTCGCACTTCATG 717
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        612 TCTGACTGGAAGAACGCCTGTGCCATGCTGAAGGACGGGACTGCAGGCTCGCACTTCATG 671

Qy        718 GCCTCTCCGCAGTGTGTGGGCTACAGCCGCAGCACAGCGGCCCCCCTCACGATGACTATG 777
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        672 GCCTCTCCGCAGTGTGTGGGCTACAGCCGCAGCACAGCGGCCCCCCTCACGATGACTATG 731

Qy        778 TGCCTGCCTGACCTGAAGGAGATCCAGAGGGCTGTGAAGCTCTGGGTGAGGTCGCTGGAT 837
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        732 TGCCTGCCTGACCTGAAGGAGATCCAGAGGGCTGTGAAGCTCTGGGTGAGGTCGCTGGAT 791

Qy        838 GCCCAGTCGGTCTACGTTGCTACTGATTCCGAGAGTTATGTGCCTGAGCTCCAACAGCTC 897
              |||||||||||||||||||||||||||||||||||||||||||||||||    | |||||||
Db        792 GCCCAGTCGGTCTACGTTGCTACTGATTCCGAGAGTTATGTGCCTGAAGCTCCACAGCTC 851

Qy        898 TTCAAA--GGGAAGGTGAAGGTGGTG--AGCCTGAAGCCTGAGGTGGCCCAGGTCGACCT 953
              || |||  || | || |||||||||| | |   || |   ||| ||||||||| ||||
Db        852 TTTAAAAGGGAAGGGGGAAGGTGGTGGAACCTGAAACCTGTAGGGGGCCCAGGTCCACCT 911

Qy        954 GTACATCCTCGGCCAAGCCGACCACTTT-ATTGGCAACTGTGTCTCCTCCTTCACTG--- 1009
              ||| || ||  | |||||||| | |||| ||||| |||||| | ||  | |  | |
Db        912 GTAAATTCTTCGGCAAGCCGAACCCTTTAATTGGGAACTGTGGCCCCCTCCTTCCCGGGC 971

Qy       1010 CCTTTGTGAAGCGGGAGCGGGACCTCCAGGGGAGGCCGTCTTCTTT 1055
              ||||||  ||  || | |||||||   |   || |   | |  ||||
Db        972 CCTTTGGAAAAGGGAAACGGGACCCTCTCAGGGGAGGGGCGGCTTT 1017


RESULT 8
BX345989
LOCUS       BX345989                841 bp    mRNA    linear   EST 23-APR-2004
DEFINITION  BX345989 Homo sapiens T CELLS (JURKAT CELL LINE) COT 10-NORMALIZED
            Homo sapiens cDNA clone CS0DJ008YF12 5-PRIME, mRNA sequence.
ACCESSION   BX345989
VERSION     BX345989.2  GI:46549776
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 841)
  AUTHORS   Li,W.B., Gruber,C., Jessee,J. and Polayes,D.
  TITLE     Full-length cDNA libraries and normalization
  JOURNAL   Unpublished (2001)
COMMENT     On May 5, 2003 this sequence version replaced gi:30377055.
            Contact: Genoscope
            Genoscope - Centre National de Sequencage
            2 rue Gaston Cremieux, CP 5706 - 91057 EVRY cedex - FRANCE
            Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr
            1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime
            end enriched, double-strand cDNA was digested with Not I and cloned
            into the Not I and EcoR V sites of the pCMVSPORT 6 vector. Library
            was normalized. Library was constructed by Life Technologies, a
            division of Invitrogen. This sequence belongs to sequence cluster
            9980.f
            For more information about this cluster, see
            http://www.genoscope.cns.fr/cdna?s=CS0AJ008DC06QP1&c=9980.f.
FEATURES             Location/Qualifiers
     source          1. .841
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="CS0DJ008YF12"
                     /cell_type="T CELLS (JURKAT CELL LINE) COT 10-NORMALIZED"
                     /cell_line="JURKAT"
                     /clone_lib="Homo sapiens T CELLS (JURKAT CELL LINE) COT
                     10-NORMALIZED"
                     /note="1st strand cDNA was primed with a NotI-oligo(dT)
                     primer. Five prime end enriched, double-strand cDNA was
                     digested with Not I and cloned into the Not I and EcoR V
                     sites of the pCMVSPORT 6 vector. Library was normalized."
ORIGIN

  Query Match           47.5%;  Score 718.4;  DB 5;  Length 841;
  Best Local Similarity  99.5%;  Pred. No. 6.9e-188;
  Matches  730;  Conservative    1;  Mismatches    2;  Indels    1;  Gaps    1;

Qy          1 ATGCCCGCGGGGCTCCTGGGACCCGGCCGGTTACCTGCTCTACTGCCCCTGCATGGGGCGC 60
              ||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        109 ATGCCTGCGGGCTCCTGGGACCCGGCCGGTTACCTGCTCTACTGCCCCTGCATGGGGCGC 168

Qy         61 TTTGGGAACCAGGCCGATCACTTCTTGGGCTCTCTGGCATTTGCAAAGCTGCTAAACCGT 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        169 TTTGGGAACCAGGCCGATCACTTCTTGGGCTCTCTGGCATTTGCAAAGCTGCTAAACCGT 228

Qy        121 ACCTTGGCTGTCCCTCCTTGGATTGAGTACCAGCATCACAAGCCTCCTTTCACCAACCTC 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        229 ACCTTGGCTGTCCCTCCTTGGATTGAGTACCAGCATCACAAGCCTCCTTTCACCAACCTC 288

Qy        181 CATGTGTCCTACCAGAAGTACTTCAAGCTGGAGCCCCTCCAGGCTTACCATCGGGTCATC 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        289 CATGTGTCCTACCAGAAGTACTTCAAGCTGGAGCCCCTCCAGGCTTACCATCGGGTCATC 348

Qy        241 AGCTTGGAGGATTTCATGGAGAAGCTGGCACCCACCCACTGGCCCCCTGAGAAGCGGGTG 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        349 AGCTTGGAGGATTTCATGGAGAAGCTGGCACCCACCCACTGGCCCCCTGAGAAGCGGGTG 408

Qy        301 GCATACTGCTTTGAGGTGGCAGCCCAGCGAAGCCCAGATAAGAAGACGTGCCCCATGAAG 360
              |||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||||
Db        409 GCATACTGCTTTGAGGTGGCAGCCCAGCGAA-CCCAGATAAGAAGACGTGCCCCATGAAG 467

Qy        361 GAAGGAAACCCCTTTGGCCCATTCTGGGATCAGTTTCATGTGAGTTTCAACAAGTCGGAG 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        468 GAAGGAAACCCCTTTGGCCCATTCTGGGATCAGTTTCATGTGAGTTTCAACAAGTCGGAG 527

Qy        421 CTTTTTACAGGCATTTCCTTCAGTGCTTCCTACAGAGAACAATGGAGCCAGAGATTTTCT 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        528 CTTTTTACAGGCATTTCCTTCAGTGCTTCCTACAGAGAACAATGGAGCCAGAGATTTTCT 587

Qy        481 CCAAAGGAACATCCGGTGCTTGCCCTGCCAGGAGCCCCAGCCCAGTTCCCCGTCCTAGAA 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        588 CCAAAGGAACATCCGGTGCTTGCCCTGCCAGGAGCCCCAGCCCAGTTCCCCGTCCTAGAG 647

Qy        541 GAACACAGGCCACTACAGAAGTACATGGTATGGTCAGACGAAATGGTGAAGACGGGAGAG 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        648 GAACACAGGCCACTACAGAAGTACATGGTATGGTCAGACGAAATGGTGAAGACGGGAGAG 707
```

```
Qy        601 GCCCAGATTCATGCCCACCTTGTCCGGCCCTATGTGGGCATTCATCTGCGCATTGGCTCT 660
              |:||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        708 GSCCAGATTCATGCCCACCTTGTCCGGCCCTATGTGGGCATTCATCTGCGCATTGGCTCT 767

Qy        661 GACTGGAAGAACGCCTGTGCCATGCTGAAGGACGGGACTGCAGGCTCGCACTTCATGGCC 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        768 GACTGGAAGAACGCCTGTGCCATGCTGAAGGACGGGACTGCAGGCTCGCACTTCATGGCC 827

Qy        721 TCTCCGCAGTGTGT 734
              ||||||||||||||
Db        828 TCTCCGCAGTGTGT 841


RESULT 9
AL135434
LOCUS       AL135434                719 bp    mRNA    linear   EST 04-SEP-2003
DEFINITION  DKFZp762O0115_r1 762 (synonym: hmel2) Homo sapiens cDNA clone
            DKFZp762O0115 5', mRNA sequence.
ACCESSION   AL135434
VERSION     AL135434.1  GI:6603621
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 719)
  AUTHORS   Ansorge,W., Wirkner,U., Mewes,W., Weil,B. and Wiemann,S.
  TITLE     EST (Ansorge,W., Wirkner,U., Mewes,H.W., Weil,B. and Wiemann,S.)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: MIPS
            MIPS
            Ingolstaedter Landstr.1, D-85764 Neuherberg, Germany
            This is the 5' sequence of the clone insert
            Clone from S. Wiemann, Molecular Genome Analysis, German Cancer
            Research Center (DKFZ); Email s.wiemann@dkfz- heidelberg.de;
            sequenced by EMBL (European Molecular Biology Laboratories,
            Heidelberg/Germany) within the cDNA sequencing consortium of the
            German Genome Project.
            No s1 sequence available.
            This clone (DKFZp762O0115) is available at the RZPD in Berlin.
            Please contact the RZPD: Ressourcenzentrum, Heubnerweg 6, 14059
            Berlin-Charlottenburg, GERMANY; Email: clone@rzpd.de.
FEATURES             Location/Qualifiers
     source          1. .719
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="DKFZp762O0115"
                     /tissue_type="melanoma (MeWo cell line)"
                     /dev_stage="adult"
                     /lab_host="DH10B"
                     /clone_lib="762 (synonym: hmel2)"
                     /note="Vector: pSport1; Site_1: NotI; Site_2: SalI"
ORIGIN

  Query Match             47.4%;  Score 717.4;  DB 1;  Length 719;
  Best Local Similarity   99.9%;  Pred. No. 1.2e-187;
  Matches  718;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy        360 GGAAGGAAACCCCTTTGGCCCATTCTGGGATCAGTTTCATGTGAGTTTCAACAAGTCGGA 419
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 GGAAGGAAACCCCTTTGGCCCATTCTGGGATCAGTTTCATGTGAGTTTCAACAAGTCGGA 60

Qy        420 GCTTTTTACAGGCATTTCCTTCAGTGCTTCCTACAGAGAACAATGGAGCCAGAGATTTTC 479
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GCTTTTTACAGGCATTTCCTTCAGTGCTTCCTACAGAGAACAATGGAGCCAGAGATTTTC 120

Qy        480 TCCAAAGGAACATCCGGTGCTTGCCCTGCCAGGAGCCCCAGCCCAGTTCCCCGTCCTAGA 539
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 TCCAAAGGAACATCCGGTGCTTGCCCTGCCAGGAGCCCCAGCCCAGTTCCCCGTCCTAGA 180

Qy        540 AGAACACAGGCCACTACAGAAGTACATGGTATGGTCAGACGAAATGGTGAAGACGGGAGA 599
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 GGAACACAGGCCACTACAGAAGTACATGGTATGGTCAGACGAAATGGTGAAGACGGGAGA 240

Qy        600 GGCCCAGATTCATGCCCACCTTGTCCGGCCCTATGTGGGCATTCATCTGCGCATTGGCTC 659
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 GGCCCAGATTCATGCCCACCTTGTCCGGCCCTATGTGGGCATTCATCTGCGCATTGGCTC 300

Qy        660 TGACTGGAAGAACGCCTGTGCCATGCTGAAGGACGGGACTGCAGGCTCGCACTTCATGGC 719
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 TGACTGGAAGAACGCCTGTGCCATGCTGAAGGACGGGACTGCAGGCTCGCACTTCATGGC 360

Qy        720 CTCTCCGCAGTGTGTGGGCTACAGCCGCAGCACAGCGGCCCCCCTCACGATGACTATGTG 779
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 CTCTCCGCAGTGTGTGGGCTACAGCCGCAGCACAGCGGCCCCCCTCACGATGACTATGTG 420

Qy        780 CCTGCCTGACCTGAAGGAGATCCAGAGGGCTGTGAAGCTCTGGGTGAGGTCGCTGGATGC 839
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 CCTGCCTGACCTGAAGGAGATCCAGAGGGCTGTGAAGCTCTGGGTGAGGTCGCTGGATGC 480

Qy        840 CCAGTCGGTCTACGTTGCTACTGATTCCGAGAGTTATGTGCCTGAGCTCCAACAGCTCTT 899
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 CCAGTCGGTCTACGTTGCTACTGATTCCGAGAGTTATGTGCCTGAGCTCCAACAGCTCTT 540

Qy        900 CAAAGGGAAGGTGAAGGTGGTGAGCCTGAAGCCTGAGGTGGCCCAGGTCGACCTGTACAT 959
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 CAAAGGGAAGGTGAAGGTGGTGAGCCTGAAGCCTGAGGTGGCCCAGGTCGACCTGTACAT 600

Qy        960 CCTCGGCCAAGCCGACCACTTTATTGGCAACTGTGTCTCCTCCTTCACTGCCTTTGTGAA 1019
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 CCTCGGCCAAGCCGACCACTTTATTGGCAACTGTGTCTCCTCCTTCACTGCCTTTGTGAA 660

Qy       1020 GCGGGAGCGGGACCTCCAGGGGAGGCCGTCTTCTTTCTTCGGCATGGACAGGCCCCCTA 1078
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 GCGGGAGCGGGACCTCCAGGGGAGGCCGTCTTCTTTCTTCGGCATGGACAGGCCCCCTA 719


RESULT 10
BI524031
LOCUS       BI524031                889 bp    mRNA    linear   EST 29-AUG-2001
DEFINITION  603052202F1 NIH_MGC_122 Homo sapiens cDNA clone IMAGE:5201616 5',
            mRNA sequence.
ACCESSION   BI524031
VERSION     BI524031.1  GI:15348823
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 889)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Life Technologies, Inc.
             cDNA Library Preparation: Life Technologies, Inc.
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM11505  row: d  column: 01
            High quality sequence stop: 880.
FEATURES             Location/Qualifiers
     source          1. .889
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:5201616"
                     /lab_host="DH10B"
                     /clone_lib="NIH_MGC_122"
                     /note="Organ: pooled lung and spleen; Vector: pCMV-SPORT6;
                     Site_1: NotI; Site_2: EcoRV (destroyed); RNA source
```

anonymous pool of 24 week female lung, 16 week female spleen, and 20-22 week male spleens. Library is oligo-dT primed and directionally cloned (EcoRV site is destroyed upon cloning). Average insert size 1.4 kb, insert size range 1-3 kb. Library is normalized and enriched for full-length clones and was constructed by C. Gruber (Invitrogen). Research Genetics tracking code 026. Note: this is a NIH_MGC Library."

ORIGIN

Query Match 47.0%; Score 712.2; DB 4; Length 889;
Best Local Similarity 99.1%; Pred. No. 3.7e-186;
Matches 758; Conservative 0; Mismatches 3; Indels 4; Gaps 4;

```
Qy          1 ATGCCCGCGGGCTCCTGGGACCCGGCCGGTTACCTGCTCTACTGCCCCTGCATGGGGCGC 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         87 ATGCCCGCGGGCTCCTGGGACCCGGCCGGTTACCTGCTCTACTGCCCCTGCATGGGGCGC 146

Qy         61 TTTGGGAACCAGGCCGATCACTTCTTGGGCTCTCTGGCATTTGCAAAGCTGCTAAACCGT 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        147 TTTGGGAACCAGGCCGATCACTTCTTGGGCTCTCTGGCATTTGCAAAGCTGCTAAACCGT 206

Qy        121 ACCTTGGCTGTCCCTCCTTGGATTGAGTACCAGCATCACAAGCCTCCTTTCACCAACCTC 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        207 ACCTTGGCTGTCCCTCCTTGGATTGAGTACCAGCATCACAAGCCTCCTTTCACCAACCTC 266

Qy        181 CATGTGTCCTACCAGAAGTACTTCAAGCTGGAGCCCCTCCAGGCTTACCATCGGGTCATC 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        267 CATGTGTCCTACCAGAAGTACTTCAAGCTGGAGCCCCTCCAGGCTTACCATCGGGTCATC 326

Qy        241 AGCTTGGAGGATTTCATGGAGAAGCTGGCACCCACCCACTGGCCCCCTGAGAAGCGGGTG 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        327 AGCTTGGAGGATTTCATGGAGAAGCTGGCACCCACCCACTGGCCCCCTGAGAAGCGGGTG 386

Qy        301 GCATACTGCTTTGAGGTGGCAGCCCAGCGAAGCCCAGATAAGAAGACGTGCCCCATGAAG 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        387 GCATACTGCTTTGAGGTGGCAGCCCAGCGAAGCCCAGATAAGAAGACGTGCCCCATGAAG 446

Qy        361 GAAGGAAACCCCTTTGGCCCATTCTGGGATCAGTTTCATGTGAGTTTCAACAAGTCGGAG 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        447 GAAGGAAACCCCTTTGGCCCATTCTGGGATCAGTTTCATGTGAGTTTCAACAAGTCGGAG 506

Qy        421 CTTTTTACAGGCATTTCCTTCAGTGCTTCCTACAGAGAACAATGGAGCCAGAGATTTTCT 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        507 CTTTTTACAGGCATTTCCTTCAGTGCTTCCTACAGAGAACAATGGAGCCAGAGATTTTCT 566

Qy        481 CCAAAGGAACATCCGGTGCTTGCCCTGCCAGGAGCCCCAGCCCAGTTCCCCGTCCTAGAA 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        567 CCAAAGGAACATCCGGTGCTTGCCCTGCCAGGAGCCCCAGCCCAGTTCCCCGTCCTAGAG 626

Qy        541 GAACACAGGCCACTACAGAAGTACATGGTATGGTCAGACGAAATGGTGAAGACGGGAGAG 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        627 GAACACAGGCCACTACAGAAGTACATGGTATGGTCAGACGAAATGGTGAAGACGGGAGAG 686

Qy        601 GCCCAGATTCATGCCCACCTTGTCCGGCC-CTATGTGGGCATTCATCTGCGCATTGGCTC 659
              |||||||||||||||| || |||||||| |||||||||||||||||||||||||||||||
Db        687 GCCCAGATTCATGCCCAGCTGGTCCGGCCACTATGTGGGCATTCATCTGCGCATTGGCTC 746

Qy        660 TGACT-GGAAGAACGCCTGTGCCATGCTGAAGGACGGG-ACTGCAGGCTCGCACTTCATG 717
              ||||| |||||||||||||||||||||||||||||||| |||||||||||||||||||||
Db        747 TGACTGGGAAGAACGCCTGTGCCATGCTGAAGGACGGGAACTGCAGGCTCGCACTTCATG 806

Qy        718 GCCTCTCCGCAGTGTGTGGGCTACAGCCGCAGCACAGCGGCCCCC 762
              |||||||||||||||||||||||| ||||||||||||||||||
Db        807 GCCTCTCCGCAGTGTGTGGGCTACAG-CGCAGCACAGCGGCCCCC 850
```

RESULT 11
BX449172
LOCUS BX449172 898 bp mRNA linear EST 06-MAY-2004
DEFINITION BX449172 Homo sapiens FETAL LIVER Homo sapiens cDNA clone CS0DM014YK11 5-PRIME, mRNA sequence.
ACCESSION BX449172
VERSION BX449172.2 GI:47067286
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 898)
AUTHORS Li,W.B., Gruber,C., Jessee,J. and Polayes,D.
TITLE Full-length cDNA libraries and normalization
JOURNAL Unpublished (2001)
COMMENT On May 22, 2003 this sequence version replaced gi:31030014.
Contact: Genoscope
Genoscope - Centre National de Sequencage
2 rue Gaston Cremieux, CP 5706 - 91057 EVRY cedex - FRANCE
Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr
1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime end enriched, double-strand cDNA was digested with Not I and cloned into the Not I and EcoRV sites of the pCMVSPORT 6 vector. Library was not normalized. Library was constructed by Life Technologies, a division of Invitrogen.
This sequence belongs to sequence cluster 9980.f
For more information about this cluster, see
http://www.genoscope.cns.fr/cdna?s=CS0AM014AF06QP1&c=9980.f.
FEATURES Location/Qualifiers
source 1. .898
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="CS0DM014YK11"
/tissue_type="FETAL LIVER"
/dev_stage="fetal"
/clone_lib="Homo sapiens FETAL LIVER"
/note="Organ: liver; Vector: pCMVSPORT_6; 1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime end enriched, double-strand cDNA was digested with Not I and cloned into the Not I and EcoRV sites of the pCMVSPORT 6 vector. Library was not normalized."
ORIGIN

Query Match 46.7%; Score 706.6; DB 5; Length 898;
Best Local Similarity 97.8%; Pred. No. 1.3e-184;
Matches 744; Conservative 4; Mismatches 10; Indels 3; Gaps 3;

```
Qy          1 ATGCCCGCGGGCTCCTGGGACCCGGCCGGTTACCTGCTCTACTGCCCCTGCATGGGGCGC 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        139 ATGCCCGCGGGCTCCTGGGACCCGGCCGGTTACCTGCTCTACTGCCCCTGCATGGGGCGC 198

Qy         61 TTTGGGAACCAGGCCGATCACTTCTTGGGCTCTCTGGCATTTGCAAAGCTGCTAAACCGT 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        199 TTTGGGAACCAGGCCGATCACTTCTTGGGCTCTCTGGCATTTGCAAAGCTGCTAAACCGT 258

Qy        121 ACCTTGGCTGTCCCTCCTTGGATTGAGTACCAGCATCACAAGCCTCCTTTCACCAACCTC 180
              ||||||||||||||||:|||||||||||||||||||||||||||||||||||||||||||
Db        259 ACCTTGGCTGTCCCTCHTTGGATTGAGTACCAGCATCACAAGCCTCCTTTCACCAACCTC 318

Qy        181 CATGTGTCCTACCAGAAGTACTTCAAGCTGGAGCCCCTCCAGGCTTACCATCGGGTCATC 240
              ||||||||||||||||||||||:|||||||||||||||||||||||||||||||||||||
Db        319 CATGTGTCCTACCAGAAGTACTTYAAGCTGGAGCCCCTCCAGGCTTACCATCGGGTCATC 378

Qy        241 AGCTTGGAGGATTTCATGGAGAAGCTGGCACCCACCCACTGGCCCCCTGAGAAGCGGGTG 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        379 AGCTTGGAGGATTTCATGGAGAAGCTGGCACCCACCCACTGGCCCCCTGAGAAGCGGGTG 438

Qy        301 GCATACTGCTTTGAGGTGGCAGCCCAGCGAAGCCCAGATAAGAAGACGTGCCCCATGAAG 360
              |||||||||||||||||||||||||||||||| |||||||||||||||||||||||||||
Db        439 GCATACTGCTTTGAGGTGGCAGCCCAGCGAA-CCCAGATAAGAAGACGTGCCCCATGAAG 497

Qy        361 GAAGGAAACCCCTTTGGCCCATTCTGGGATCAGTTTCATGTGAGTTTCAACAAGTCGGAG 420
              |||||||||||||||||||||||||||||||||||||||||||||| ||| ||||||||||
Db        498 GAAGGAAACCCCTTTGGCCCATTCTGGGATCAGTTTCATGTAAGTGTCAACAAGTCGGAG 557

Qy        421 CTTTTTACAGGCATTTCCTTCAGTGCTTCCTACAGAGAACAATGGAGCCAGAGATTTTCT 480
              |||||||||||||||||||||||||||||| |||||||||||||||||||||||||||||
```

```
Db        558 CTTTTTACAGGCATTTCCTTCAGTGCTTCGTACAGAGAACAATGGAGCCAGAGATTTTCT 617

Qy        481 CCAAAGGAACATCCGGTGCTTGCCCTGCCAGGAGCCCCAGCCCAGTTCCCCGTCCTAGAA 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        618 CCAAAGGAACATCCGGTGCTTGCCCTGCCAGGAGCCCCAGCCCAGTTCCCCGTCCTAGAG 677

Qy        541 GAACACAGGCCACTACAGAAGTACATGGTATGGTCAGACG-AAATGGTGAAGACGGGAGA 599
              |||||||||||||||||||||||||||||||||||||||| ||||||||||||||||||
Db        678 GAACACAGGCCACTACAGAAGTACATGGTATGGTCAGACGAAAATGGTGAAGACGGGAGA 737

Qy        600 GGCCCAGATTCATGCCCACCTTGTCCGGCCCTATGTGGGCATTCATCTGCGCATTGGCTC 659
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        738 CGCCCAGATTCATGCCCACCTTGTCCGGCCCTATGTGGGCATTCATCTGCGCATTGGCTC 797

Qy        660 TGACTGGAAGAACGCCTGTGCCATGCTGAAGGACGGGACTGCAGGCTCGCACTTCATGGC 719
              |||||||||||||||||| | ||||||||||||||||||||||||||||||| ||||
Db        798 TGACTGGAAGAACGCCTGTCCAATGCTGAAGGACGGGACTGCAGGCTCGCACTTAATGGG 857

Qy        720 CT-CTCCGCAGTGTGTGGGCTACAGCCGCAGCACAGCGGCC 759
              || |||||||||||||||:|||||||||||||| :||||
Db        858 CTYCTCCGCAGTGTGTGGGCDACAGCCGCAGCACACSGGCC 898
```

RESULT 12
BM475212

LOCUS	BM475212 704 bp mRNA linear EST 05-FEB-2002
DEFINITION	AGENCOURT_6477457 NIH_MGC_88 Homo sapiens cDNA clone IMAGE:5559347 5', mRNA sequence.
ACCESSION	BM475212
VERSION	BM475212.1 GI:18524254
KEYWORDS	EST.
SOURCE	Homo sapiens (human)
ORGANISM	Homo sapiens Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE	1 (bases 1 to 704)
AUTHORS	NIH-MGC http://mgc.nci.nih.gov/.
TITLE	National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL	Unpublished (1999)
COMMENT	Contact: Robert Strausberg, Ph.D. Email: cgapbs-r@mail.nih.gov Tissue Procurement: ATCC cDNA Library Preparation: Life Technologies, Inc. cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL) DNA Sequencing by: Agencourt Bioscience Corporation Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov Plate: LLAM12283 row: m column: 12 High quality sequence stop: 685.

FEATURES Location/Qualifiers

```
     source          1. .704
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:5559347"
                     /tissue_type="duodenal adenocarcinoma, cell line"
                     /lab_host="DH10B (phage-resistant)"
                     /clone_lib="NIH_MGC_88"
                     /note="Organ: small intestine; Vector: pCMV-SPORT6;
                     Site_1: NotI; Site_2: SalI; Cloned unidirectionally;
                     oligo-dT primed. Average insert size 1.767 kb. Library
                     enriched for full-length clones and constructed by Life
                     Technologies. Note: this is a NIH_MGC Library."
ORIGIN

  Query Match            45.2%;  Score 683.8;  DB 4;  Length 704;
  Best Local Similarity  99.7%;  Pred. No. 2.7e-178;
  Matches  685;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy        797 AGATCCAGAGGGCTGTGAAGCTCTGGGTGAGGTCGCTGGATGCCCAGTCGGTCTACGTTG 856
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          3 AGATCCAGAGGGCTGTGAAGCTCTGGGTGAGGTCGCTGGATGCCCAGTCGGTCTACGTTG 62

Qy        857 CTACTGATTCCGAGAGTTATGTGCCTGAGCTCCAACAGCTCTTCAAAGGGAAGGTGAAGG 916
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         63 CTACTGATTCCGAGAGTTATGTGCCTGAGCTCCAACAGCTCTTCAAAGGGAAGGTGAAGG 122

Qy        917 TGGTGAGCCTGAAGCCTGAGGTGGCCCAGGTCGACCTGTACATCCTCGGCCAAGCCGACC 976
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        123 TGGTGAGCCTGAAGCCTGAGGTGGCCCAGGTCGACCTGTACATCCTCGGCCAAGCCGACC 182

Qy        977 ACTTTATTGGCAACTGTGTCTCCTCCTTCACTGCCTTTGTGAAGCGGGAGCGGGACCTCC 1036
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        183 ACTTTATTGGCAACTGTGTCTCCTCCTTCACTGCCTTTGTGAAGCGGGAGCGGGACCTCC 242

Qy       1037 AGGGGAGGCCGTCTTCTTTCTTCGGCATGGACAGGCCCCCTAAGCTGCGGGACGAGTTCT 1096
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        243 AGGGGAGGCCGTCTTCTTTCTTCGGCATGGACAGGCCCCCTAAGCTGCGGGACGAGTTCT 302

Qy       1097 GATTCTGGCCGGAGCACCAGACCCTCTGATCCTGGAGGGACCAGAGTCTGAGCTGGTCCT 1156
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        303 GATTCTGGCCGGAGCACCAGACCCTCTGATCCTGGAGGGACCAGAGTCTGAGCTGGTCCT 362

Qy       1157 TCCAGCCAGGCCTGGCAGCCAGAGGTGCTCCGGGATTGCAAACTCCTCTTCTCACCTGCC 1216
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        363 TCCAGCCAGGCCTGGCAGCCAGAGGTGCTCCGGGATTGCAGACTCCTCTTCTCACCTGCC 422

Qy       1217 AAAGATGGAGAAGAGTGCCAGGGACCCCTCAAGGAGGGAGACGCTCCATATCCCAGGGCA 1276
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        423 AAAGATGGAGAAGAGTGCCAGGGACCCCTCAAGGAGGGAGACGCTCCATATCCCAGGGCA 482

Qy       1277 TAGGACTTGCAGGTTCCTAGGAGCAGGAGCATCTCCCATCGCACGTGCTTTCTGCTCTTC 1336
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        483 TAGGACTTGCAGGTTCCTAGGAGCAGGAGCATCTCCCATCGCACGTGCTTTCTGCTCTTC 542

Qy       1337 TGGGAATTTCTCACACTGGCAAAGCAGTCCAGCCTCCGTCTTCTGGTCCACTCTGCTCTG 1396
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        543 TGGGAATTTCTCACACTGGCAAAGCAGTCCAGCCTCCGTCTTCTGGTCCACTCTGCTCTG 602

Qy       1397 AGCAGCCTGGGATGCTGAACTCTTCAGAGAGATTTTTTTATAGAGAGATTTCTATAATTT 1456
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        603 AGCAGCCTGGGATGCTGAACTCTTCAGAGAGATTTTTTTATAGAGAGATTTCTATAATTT 662

Qy       1457 TGATACAAGGTCATGACTATCCTAGAA 1483
              |||||||||||||||||||||||| ||
Db        663 TGATACAAGGTCATGACTATCCTAAAA 689
```

RESULT 13
CO774833

LOCUS	CO774833 942 bp mRNA linear EST 04-AUG-2004
DEFINITION	ILLUMIGEN_MCQ_51950 Katze_MNLV Macaca nemestrina cDNA clone IBIUW:29960 5' similar to Bases 7 to 852 highly similar to human POFUT1 (Hs.178292), mRNA sequence.
ACCESSION	CO774833
VERSION	CO774833.1 GI:50976096
KEYWORDS	EST.
SOURCE	Macaca nemestrina (pig-tailed macaque)
ORGANISM	Macaca nemestrina Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Cercopithecidae; Cercopithecinae; Macaca.
REFERENCE	1 (bases 1 to 942)
AUTHORS	Katze,M.G., Thomas,M., Korth,M., Iadonato,S.P. and Magness,C.L.
TITLE	Large-scale Rhesus Macaque cDNA Sequencing
JOURNAL	Unpublished (2003)
COMMENT	Contact: C. Magness Illumigen Biosciences Inc. 2203 Airport Way S, Suite 450, Seattle, WA 98134, USA Tel: 2063780400 Fax: 2063780408 Email: cmagness@illumigen.com Sequenced on 2004.07.21. 611 Q20 bases. Library Preparation: Prof. Michael Katze Lab at University of Washington DNA Sequencing: Illumigen Biosciences Inc. For further information, see

```
            http://www.macaque.org
            PCR PRimers
            FORWARD: CCCTCACTAAAGGGAACAAAA
            BACKWARD: CACTATAGGGCGAATTGGGTA
            Insert Length: 942   Std Error: 0.00
            Plate: CL000531  row: E  column: 09
            Seq primer: CCCTCACTAAAGGGAACAAAA
            POLYA=No.
FEATURES              Location/Qualifiers
     source           1. .942
                      /organism="Macaca nemestrina"
                      /mol_type="mRNA"
                      /strain="Indian"
                      /db_xref="taxon:9545"
                      /clone="IBIUW:29960"
                      /sex="male"
                      /lab_host="Electromax DH10B"
                      /clone_lib="Katze_MNLV"
                      /note="Organ: liver; Vector: pDONR 222; Site_1: BsrG I;
                      Site_2: BsrG I; Created from CloneMiner cDNA Library
                      Construction kit (catalog #18249-029)"
ORIGIN

  Query Match             44.5%;  Score 673.2;  DB 7;  Length 942;
  Best Local Similarity   94.1%;  Pred. No. 2.6e-175;
  Matches  743;  Conservative    0;  Mismatches   43;  Indels    4;  Gaps    4;

Qy          1 ATGCCCGCGGGCTCCTGGGACCCGGCCGGTTACCTGCTCTACTGCCCCTGCATGGGGCGC 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         95 ATGCCCGCGGGCTCCTGGGACCCGGCCGGTTACCTGCTCTACTGCCCCTGCATGGGGCGC 154

Qy         61 TTTGGGAACCAGGCCGATCACTTCTTGGGCTCTCTGGCATTTGCAAAGCTGCTAAACCGT 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        155 TTTGGGAACCAGGCCGATCACTTCTTGGGCTCTCTGGCATTTGCAAAGCTGCTAAACCGC 214

Qy        121 ACCTTGGCTGTCCCTCCTTGGATTGAGTACCAGCATCACAAGCCTCCTTTCACCAACCTC 180
              |||||||||||||||||||||||||||||||||||||||||||||| ||||||||||||
Db        215 ACCTTGGCTGTCCCTCCTTGGATTGAGTACCAGCATCACAAGCCTCCCTTCACCAACCTC 274

Qy        181 CATGTGTCCTACCAGAAGTACTTCAAGCTGGAGCCCCTCCAGGCTTACCATCGGGTCATC 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        275 CATGTGTCCTACCAGAAGTACTTCAAGCTGGAGCCCCTCCAGGCTTACCATCGGGTCATC 334

Qy        241 AGCTTGGAGGATTTCATGGAGAAGCTGGCACCCACCCACTGGCCCCCTGAGAAGCGGGTG 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        335 AGCTTGGAGGATTTCATGGAGAAGCTGGCACCCACCCACTGGCCCCCTGAGAAGCGGGTG 394

Qy        301 GCATACTGCTTTGAGGTGGCAGCCCAGCGAAGCCCAGATAAGAAGACGTGCCCCATGAAG 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        395 GCATACTGCTTTGAGGTGGCAGCCCAGCGAAGCCCAGATAAGAAGACGTGCCCCATGAAG 454

Qy        361 GAAGGAAACCCCTTTGGCCCATTCTGGGATCAGTTTCATGTGAGTTTCAACAAGTCGGAG 420
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        455 GAAGGAAACCCCTTTGGCCCATTCTGGGATCAGTTTCATGTGAGTTTCAACAAGTCTGAG 514

Qy        421 CTTTTTACAGGCATTTCCTTCAGTGCTTCCTACAGAGAACAATGGAGCCAGAGATTTTCT 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        515 CTTTTTACAGGCATTTCCTTCAGTGCTTCCTACAGAGAACAATGGAGCCAGAGATTTTCT 574

Qy        481 CCAAAGGAACATCCGGTGCTTGCCCTGCCAGGAGCCCCAGCCCAGTTCCCCGTCCTAGAA 540
              |||||||||||||||||||||||||||||||| |||||||| |||||||||| ||||| |||
Db        575 CCAAAGGAACATCCGGTGCTTGCCCTGCCGGGAGCCCCGGCCCAGTTCCCTGTCCTGGAA 634

Qy        541 GAACACAGGCCACTACAGAAGTACATGGTATGGTCAGACGAAATGGTGAAGACGGGAGAG 600
              |||||||| ||||| || |||||||||||||||||||| ||||||||| |||||||||
Db        635 GAACACAGACCACTGCAAAAGTACATGGTATGGTCAGACCAAATGGTGAGGACGGGAGAA 694

Qy        601 GCCCAGATTCATGCCCACCTTGT-CCGGCCCTATGTGGGCATTCATCTGCGCATTGGC-T 658
              ||||||| ||  |||||||||| |||||||||||||||||||| ||| || | |||| |
Db        695 GCCCAGAATCCCGCCCACCTTGTCCCGGCCCTATGTGGGCATTCCTCTACGTAATGGCTT 754

Qy        659 CTGACTGGAAGAACGCCTGTGCCATGCTGAAGGACGGGACTGCAGGCTCGCACTTCATGG 718
              |||||||||| || |||||||| ||||| |||||||| ||||   |||||   ||| |||

Db        755 CTGACTGGAAAAATGCCTGTGCTATGCTAAAGGACGGAACTGGCAGCTCGACTTTCTTGG 814

Qy        719 CCTCTCCGCAGTGTGTGGG-CTACAGCCGCAGCACAG-CGGCCCCCCTCACGATGACTAT 776
              |||||||  | |||| ||| |||| || || ||||| || |||||||| ||| |||||||
Db        815 CCTCTCCCGAATGTGGGGGCCTACGCCCCCAACACAGCCGCCCCCCCTCCCGAAGACTAT 874

Qy        777 GTGCCTGCCT 786
              ||||||||||
Db        875 GTGCCTGCCT 884


RESULT 14
BG747760
LOCUS       BG747760                 808 bp    mRNA    linear   EST 15-MAY-2001
DEFINITION  602705255F1 NIH_MGC_43 Homo sapiens cDNA clone IMAGE:4842061 5',
            mRNA sequence.
ACCESSION   BG747760
VERSION     BG747760.1  GI:14058413
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 808)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: ATCC
             cDNA Library Preparation: Ling Hong/Rubin Laboratory
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLCM1675  row: n  column: 14
            High quality sequence stop: 808.
FEATURES              Location/Qualifiers
     source           1. .808
                      /organism="Homo sapiens"
                      /mol_type="mRNA"
                      /db_xref="taxon:9606"
                      /clone="IMAGE:4842061"
                      /tissue_type="normal pigmented retinal epithelium"
                      /lab_host="DH10B (phage-resistant)"
                      /clone_lib="NIH_MGC_43"
                      /note="Organ: eye; Vector: pOTB7; Site_1: XhoI; Site_2:
                      EcoRI; cDNA made by oligo-dT priming.  Directionally
                      cloned into EcoRI/XhoI sites using the following 5'
                      adaptor: GGCACGAG(G).  Library constructed  by Ling Hong
                      in the laboratory of Gerald M. Rubin (University of
                      California, Berkeley) using ZAP-cDNA synthesis kit
                      (Stratagene) and  Superscript II RT (Life Technologies).
                      Note: this is a NIH_MGC Library. |"
ORIGIN

  Query Match             43.7%;  Score 662.2;  DB 4;  Length 808;
  Best Local Similarity   95.3%;  Pred. No. 2.8e-172;
  Matches  693;  Conservative    0;  Mismatches   33;  Indels    1;  Gaps    1;

Qy          1 ATGCCCGCGGGCTCCTGGGACCCGGCCGGTTACCTGCTCTACTGCCCCTGCATGGGGCGC 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         82 ATGCCCGCGGGCTCCTGGGACCCGGCCGGTTACCTGCTCTACTGCCCCTGCATGGGGCGC 141

Qy         61 TTTGGGAACCAGGCCGATCACTTCTTGGGCTCTCTGGCATTTGCAAAGCTGCTAAACCGT 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        142 TTTGGGAACCAGGCCGATCACTTCTTGGGCTCTCTGGCATTTGCAAAGCTGCTAAACCGT 201

Qy        121 ACCTTGGCTGTCCCTCCTTGGATTGAGTACCAGCATCACAAGCCTCCTTTCACCAACCTC 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        202 ACCTTGGCTGTCCCTCCTTGGATTGAGTACCAGCATCACAAGCCTCCTTTCACCAACCTC 261
```

```
Qy        181 CATGTGTCCTACCAGAAGTACTTCAAGCTGGAGCCCCTCCAGGCTTACCATCGGGTCATC 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        262 CATGTGTCCTACCAGAAGTACTTCAAGCTGGAGCCCCTCCAGGCTTACCATCGGGTCATC 321

Qy        241 AGCTTGGAGGATTTCATGGAGAAGCTGGCACCCACCCACTGGCCCCCTGAGAAGCGGGTG 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        322 AGCTTGGAGGATTTCATGGAGAAGCTGGCACCCACCCACTGGCCCCCTGAGAAGCGGGTG 381

Qy        301 GCATACTGCTTTGAGGTGGCAGCCCAGCGAAGCCCAGATAAGAAGACGTGCCCCATGAAG 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        382 GCATACTGCTTTGAGGTGGCAGCCCAGCGAAGCCCAGATAAGAAGACGTGCCCCATGAAG 441

Qy        361 GAAGGAAACCCCTTTGGCCCATTCTGGGATCAGTTTCATGTGAGTTTCAACAAGTCGGAG 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        442 GAAGGAAACCCCTTTGGCCCATTCTGGGATCAGTTTCATGTGAGTTTCAACAAGTCGGAG 501

Qy        421 CTTTTTACAGGCATTTCCTTCAGTGCTTCCTACAGAGAACAATGGAGCCAGAGATTTTCT 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        502 CTTTTTACAGGCATTTCCTTCAGTGCTTCCTACAGAGAACAATGGAGCCAGAGATTTTCT 561

Qy        481 CCAAAGGAACATCCGGTGCTTGCCCTGCCAGGAGCCCCAGCCCAGTTCCCCGTCCTAGAA 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        562 CCAAAGGAACATCCGGTGCTTGCCCTGCCAGGAGCCCCAGCCCAGTTCCCCGTCCTAGAG 621

Qy        541 GAACACAGGCCACTACAGAAGTACATGGTATGGTCAGACGAAATGGTGAAGAC-GGGAGA 599
              ||||||||||||||||||||||||||||||||||||||||||||||||||||| ||||||
Db        622 GAACACAGGCCACTACAGAAGTACATGGTATGGTCAGACGAAATGGTGAAGACGGGGAGA 681

Qy        600 GGCCCAGATTCATGCCCACCTTGTCCGGCCCTATGTGGGCATTCATCTGCGCATTGGCTC 659
              |||||||||||||||| || |||||| ||||||| ||   |||          | | ||
Db        682 GGCCCAGATTCATGCCCAACTGGTCCGGACCTATGTTGGGCATCAATCTGCGCATTGGTC 741

Qy        660 TGACTGGAAGAACGCCTGTGCCATGCTGAAGGACGGGACTGCAGGCTCGCACTTCATGGC 719
              |||||||||||||||||||||||||||||||||||    |||||||    |    |
Db        742 TGACTGGAAGAACGCCTGTGCCATGCTGAAGGACGGGGACTGAGGCTCGGCACTTCATGG 801

Qy        720 CTCTCCG 726
              |||||||
Db        802 CTCTCCG 808
```

RESULT 15
BP144264
LOCUS BP144264 801 bp mRNA linear EST 30-DEC-2003
DEFINITION BP144264 full-length enriched swine cDNA library, adult ovary Sus scrofa cDNA clone OVR010034G09 5', mRNA sequence.
ACCESSION BP144264
VERSION BP144264.1 GI:40393735
KEYWORDS EST.
SOURCE Sus scrofa (pig)
ORGANISM Sus scrofa
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
REFERENCE 1 (bases 1 to 801)
AUTHORS Uenishi,H., Eguchi,T., Suzuki,K., Sawazaki,T., Toki,D., Shinkai,H., Okumura,N., Hamasima,N. and Awata,T.
TITLE PEDE (Pig EST Data Explorer): construction of a database for ESTs derived from porcine full-length cDNA libraries
JOURNAL Nucleic Acids Res. 32 (1), D484-D488 (2004)
COMMENT Contact: Hirohide Uenishi
Animal Genome Laboratory, Genome Research Department
National Institute of Agrobiological Sciences
2 Ikenodai, Tsukuba, Ibaraki 305-8602, Japan
Tel: +81-29-838-8627
Fax: +81-29-838-8627
Email: huenishi@affrc.go.jp
EST project with full-length enriched cDNA libraries carried out in Animal Genome Research Program (Japan) by National Institute of Agrobiological Sciences and STAFF-Institute
Single pass sequencing of clones derived from oligo-capped cDNA library
Vector sequences were eliminated by RepeatMasker version 2002/07/13 and crossmatch version 0.990319

Low quality bases were trimmed based on the quality values.
FEATURES Location/Qualifiers
source 1. .801
/organism="Sus scrofa"
/mol_type="mRNA"
/db_xref="taxon:9823"
/clone="OVR010034G09"
/tissue_type="ovary"
/dev_stage="adult"
/clone_lib="full-length enriched swine cDNA library, adult ovary"
ORIGIN

Query Match 43.6%; Score 660.2; DB 5; Length 801;
Best Local Similarity 89.0%; Pred. No. 1e-171;
Matches 713; Conservative 0; Mismatches 88; Indels 0; Gaps 0;

```
Qy        194 AGAAGTACTTCAAGCTGGAGCCCCTCCAGGCTTACCATCGGGTCATCAGCTTGGAGGATT 253
              | ||||||||||||||||||||||||||||  ||||| |||||||||||| ||||||| |
Db          1 AAAAGTACTTCAAGCTGGAGCCCCTCCAGGTCTACCACCGGGTCATCAGCCTGGAGGACT 60

Qy        254 TCATGGAGAAGCTGGCACCCACCCACTGGCCCCCTGAGAAGCGGGTGGCATACTGCTTTG 313
              ||||||||||||||||||||||||||||||||||||||||||||||| ||||||||||||||
Db         61 TCATGGAGAAGCTGGCACCCACCCACTGGCCCCCTGAGAGGCGGGTGGCATACTGTTTTG 120

Qy        314 AGGTGGCAGCCCAGCGAAGCCCAGATAAGAAGACGTGCCCCATGAAGGAAGGAAACCCCT 373
              |||||||||| |||||||||||||||||||||||| ||||||||||||||||||||||||||
Db        121 AGGTGGCAGCACAGCGAAGCCCAGATAAGAAGACATGCCCCATGAAGGAAGGAAACCCCT 180

Qy        374 TTGGCCCATTCTGGGATCAGTTTCATGTGAGTTTCAACAAGTCGGAGCTTTTTACAGGCA 433
              |||||||||| |||||||||||||||||||||||||||||||| || |||||| ||||||
Db        181 TTGGCCCATTTTGGGATCAGTTTCATGTGAGTTTCAACAAGTCAGAACTTTTTGCAGGCA 240

Qy        434 TTTCCTTCAGTGCTTCCTACAGAGAACAATGGAGCCAGAGATTTTCTCCAAAGGAACATC 493
              |||||||||||||| ||||||| ||| || |||| ||| ||||||||||||||||||||||
Db        241 TTTCCTTCAGTGCCTCCTACAAAGACCAGTGGATCCAAAGATTTTCTCCAAAGGAACATC 300

Qy        494 CGGTGCTTGCCCTGCCAGGAGCCCCAGCCCAGTTCCCCGTCCTAGAAGAACACAGGCCAC 553
              | |||||||||||||||| ||||| || |||||||||||||| || ||||| |||||||
Db        301 CAGTGCTTGCCCTGCCAGGGGCCCCGGCTCAGTTCCCCGTCCTGGAGGAACATAGGCCAC 360

Qy        554 TACAGAAGTACATGGTATGGTCAGACGAAATGGTGAAGACGGGAGAGGCCCAGATTCATG 613
              | ||||| |||||||| |||||||||||| ||||||| ||| || ||||||||||| | ||
Db        361 TTCAGAAATACATGGTGTGGTCAGACGAGATGGTGAGGACAGGGGAGGCCCAGATCCGTG 420

Qy        614 CCCACCTTGTCCGGCCCTATGTGGGCATTCATCTGCGCATTGGCTCTGACTGGAAGAACG 673
              |||||||  |||||||||| |||||||||| ||||||||||||||||||||||||||||||
Db        421 CCCACCTCATCCGGCCCTACGTGGGCATTCACCTGCGCATTGGCTCTGACTGGAAGAACG 480

Qy        674 CCTGTGCCATGCTGAAGGACGGGACTGCAGGCTCGCACTTCATGGCCTCTCCGCAGTGTG 733
              | || |||||||||||||||| ||||| ||| | |||||||||||||| ||||||||  |
Db        481 CGTGCGCCATGCTGAAGGACGGCACTGCCGGCGCTCACTTCATGGCCTCCCCGCAGTGCG 540

Qy        734 TGGGCTACAGCCGCAGCACAGCGGCCCCCCTCACGATGACTATGTGCCTGCCTGACCTGA 793
              |||| |||||||||  ||| || ||||| ||||| ||||| |||||||| || |||||||
Db        541 TGGGTTACAGCCGCCACACCGCCGCCCCGCTCACCATGACCATGTGCCTCCCCGACCTGA 600

Qy        794 AGGAGATCCAGAGGGCTGTGAAGCTCTGGGTGAGGTCGCTGGATGCCCAGTCGGTCTACG 853
              ||||||||  | | ||  |||||||||||||| | |||||   |||||||| ||||||
Db        601 AGGAGATCAGGCGCGCCCTGAAGCTCTGGGTGACGGCGCTGAGCGCCCAGTCCGTCTACA 660

Qy        854 TTGCTACTGATTCCGAGAGTTATGTGCCTGAGCTCCAACAGCTCTTCAAAGGGAAGGTGA 913
              | || || || || |||||||  |||| ||| |||| |||||||||||||| |||||||
Db        661 TCGCCACGGACTCTGAGAGTTACCTGCCGGAGATCCAGCAGCTCTTCAAAGGAAAGGTGA 720

Qy        914 AGGTGGTGAGCCTGAAGCCTGAGGTGGCCCAGGTCGACCTGTACATCCTCGGCCAAGCCG 973
              |||||||||| ||||| |||||||||||||| |||||||||||||||| ||||| || |
Db        721 AGGTGGTGAGTCTGAAACCTGAGGTGGCCCAGATCGACCTGTACATCCTTGGCCAGGCTG 780

Qy        974 ACCACTTTATTGGCAACTGTG 994
              |||||||||||||||||||||
Db        781 ACCACTTTATTGGCAACTGTG 801
```

Search completed: November 9, 2005, 10:38:05
Job time : 5495.13 secs

GenCore version 5.1.6

OM nucleic - nucleic search, using sw model

Run on: November 8, 2005, 23:53:44 ; Search time 4962.62 Seconds
(without alignments)
10740.434 Million cell updates/sec

Title: US-09-774-954-16
Perfect score: 1100
Sequence: 1 ATGCCCGCGGGCTCCTGGGA.....CTGCGGGACGAGTTCTGATT 1100

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 4708233 seqs, 24227607955 residues

Total number of hits satisfying chosen parameters: 9416466

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl:*
1: gb_ba:*
2: gb_htg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vi:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	1100	100.0	1100	6	AR163457	AR163457 Sequence
2	1100	100.0	1100	6	BD103196	BD103196 O-fucosyl
3	1100	100.0	1514	6	AR163451	AR163451 Sequence
4	1100	100.0	1514	6	BD103188	BD103188 O-fucosyl
5	1098.4	99.9	1300	6	AR163452	AR163452 Sequence
6	1098.4	99.9	1300	6	BD103189	BD103189 O-fucosyl
7	1098.4	99.9	5249	9	AF375884	AF375884 Homo sapi
8	1098.4	99.9	11284	6	BD103190	BD103190 O-fucosyl
9	1096.8	99.7	5189	9	D80002	D80002 Homo sapien
10	1096.4	99.7	1167	6	CQ727777	CQ727777 Sequence
11	1095.8	99.6	5266	6	AX780153	AX780153 Sequence
12	1095.2	99.6	4560	9	HSM803330	AL832023 Homo sapi
13	1086.8	98.8	1167	9	AJ781500	AJ781500 Pan trogl
14	1033.4	93.9	5009	6	AR163453	AR163453 Sequence
15	1033.4	93.9	5009	6	BD103191	BD103191 O-fucosyl
16	916.2	83.3	1176	4	AY344580	AY344580 Bos tauru
17	908.2	82.6	1176	4	AJ781503	AJ781503 Bos tauru
18	901.6	82.0	1134	4	SSC567917	AJ567917 Sus scrof
19	889.6	80.9	1188	10	AJ781499	AJ781499 Rattus no
20	885.8	80.5	1302	10	AF375885	AF375885 Mus muscu
21	885.8	80.5	1529	10	BC046295	BC046295 Mus muscu
22	752.8	68.4	1056	4	AY344581	AY344581 Bos tauru
23	652.2	59.3	1732	5	GGA535754	AJ535754 Gallus ga
24	652.2	59.3	2546	5	AJ720352	AJ720352 Gallus ga
25	586.4	53.3	1502	5	BC082519	BC082519 Xenopus t
26	586.4	53.3	1538	5	AJ781498	AJ781498 Silurana
27	563.6	51.2	1155	5	AJ606070	AJ606070 Fugu rubr
28	552.8	50.3	1505	5	XLA514425	AJ514425 Xenopus l
29	539	49.0	1164	5	AJ781504	AJ781504 Tetraodon
30	538.4	48.9	1549	5	DRE579536	AJ579536 Danio rer
31	471.4	42.9	1524	9	BC000582	BC000582 Homo sapi
32	445.4	40.5	920	5	AJ719585	AJ719585 Gallus ga
33	251.2	22.8	145414	9	HSA392M18	AL121897 Human DNA
34	201.6	18.3	1841	3	AK112708	AK112708 Ciona int
35	193.4	17.6	2393	9	AK124093	AK124093 Homo sapi
c 36	188.8	17.2	240336	2	AC094844	AC094844 Rattus no
c 37	186.6	17.0	240510	2	AC134371	AC134371 Rattus no
38	184.6	16.8	200329	10	AC078911	AC078911 Mus muscu
39	184.6	16.8	208614	10	AL807380	AL807380 Mouse DNA
40	184	16.7	1146	3	AJ781502	AJ781502 Ciona sav
41	172.8	15.7	1209	3	AJ831490	AJ831490 Drosophil
42	168	15.3	1295	3	AY118651	AY118651 Drosophil
43	168	15.3	1320	6	AR509393	AR509393 Sequence
44	166.8	15.2	1209	3	AJ781501	AJ781501 Drosophil
45	166.4	15.1	1209	6	CQ590403	CQ590403 Sequence

ALIGNMENTS

RESULT 1
AR163457
LOCUS AR163457 1100 bp DNA linear PAT 17-OCT-2001
DEFINITION Sequence 16 from patent US 6270987.
ACCESSION AR163457
VERSION AR163457.1 GI:16234062
KEYWORDS .
SOURCE Unknown.
ORGANISM Unknown.
Unclassified.
REFERENCE 1 (bases 1 to 1100)
AUTHORS Wang,Y. and Spellman,M.W.
TITLE O-fucosyltransferase
JOURNAL Patent: US 6270987-A 16 07-AUG-2001;
FEATURES Location/Qualifiers
source 1. .1100
/organism="unknown"
/mol_type="unassigned DNA"
ORIGIN

Query Match 100.0%; Score 1100; DB 6; Length 1100;
Best Local Similarity 100.0%; Pred. No. 1e-253;
Matches 1100; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 ATGCCCGCGGGCTCCTGGGACCCGGCCGGTTACCTGCTCTACTGCCCCTGCATGGGGCGC 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGCCCGCGGGCTCCTGGGACCCGGCCGGTTACCTGCTCTACTGCCCCTGCATGGGGCGC 60

Qy         61 TTTGGGAACCAGGCCGATCACTTCTTGGGCTCTCTGGCATTTGCAAAGCTGCTAAACCGT 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 TTTGGGAACCAGGCCGATCACTTCTTGGGCTCTCTGGCATTTGCAAAGCTGCTAAACCGT 120

Qy        121 ACCTTGGCTGTCCCTCCTTGGATTGAGTACCAGCATCACAAGCCTCCTTTCACCAACCTC 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 ACCTTGGCTGTCCCTCCTTGGATTGAGTACCAGCATCACAAGCCTCCTTTCACCAACCTC 180

Qy        181 CATGTGTCCTACCAGAAGTACTTCAAGCTGGAGCCCCTCCAGGCTTACCATCGGGTCATC 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 CATGTGTCCTACCAGAAGTACTTCAAGCTGGAGCCCCTCCAGGCTTACCATCGGGTCATC 240

Qy        241 AGCTTGGAGGATTTCATGGAGAAGCTGGCACCCACCCACTGGCCCCCTGAGAAGCGGGTG 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db        241 AGCTTGGAGGATTTCATGGAGAAGCTGGCACCCACCCACTGGCCCCCTGAGAAGCGGGTG 300

Qy        301 GCATACTGCTTTGAGGTGGCAGCCCAGCGAAGCCCAGATAAGAAGACGTGCCCCATGAAG 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 GCATACTGCTTTGAGGTGGCAGCCCAGCGAAGCCCAGATAAGAAGACGTGCCCCATGAAG 360

Qy        361 GAAGGAAACCCCTTTGGCCCATTCTGGGATCAGTTTCATGTGAGTTTCAACAAGTCGGAG 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 GAAGGAAACCCCTTTGGCCCATTCTGGGATCAGTTTCATGTGAGTTTCAACAAGTCGGAG 420

Qy        421 CTTTTTACAGGCATTTCCTTCAGTGCTTCCTACAGAGAACAATGGAGCCAGAGATTTTCT 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 CTTTTTACAGGCATTTCCTTCAGTGCTTCCTACAGAGAACAATGGAGCCAGAGATTTTCT 480

Qy        481 CCAAAGGAACATCCGGTGCTTGCCCTGCCAGGAGCCCCAGCCCAGTTCCCCGTCCTAGAA 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 CCAAAGGAACATCCGGTGCTTGCCCTGCCAGGAGCCCCAGCCCAGTTCCCCGTCCTAGAA 540

Qy        541 GAACACAGGCCACTACAGAAGTACATGGTATGGTCAGACGAAATGGTGAAGACGGGAGAG 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 GAACACAGGCCACTACAGAAGTACATGGTATGGTCAGACGAAATGGTGAAGACGGGAGAG 600

Qy        601 GCCCAGATTCATGCCCACCTTGTCCGGCCCTATGTGGGCATTCATCTGCGCATTGGCTCT 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 GCCCAGATTCATGCCCACCTTGTCCGGCCCTATGTGGGCATTCATCTGCGCATTGGCTCT 660

Qy        661 GACTGGAAGAACGCCTGTGCCATGCTGAAGGACGGGACTGCAGGCTCGCACTTCATGGCC 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 GACTGGAAGAACGCCTGTGCCATGCTGAAGGACGGGACTGCAGGCTCGCACTTCATGGCC 720

Qy        721 TCTCCGCAGTGTGTGGGCTACAGCCGCAGCACAGCGGCCCCCCTCACGATGACTATGTGC 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 TCTCCGCAGTGTGTGGGCTACAGCCGCAGCACAGCGGCCCCCCTCACGATGACTATGTGC 780

Qy        781 CTGCCTGACCTGAAGGAGATCCAGAGGGCTGTGAAGCTCTGGGTGAGGTCGCTGGATGCC 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 CTGCCTGACCTGAAGGAGATCCAGAGGGCTGTGAAGCTCTGGGTGAGGTCGCTGGATGCC 840

Qy        841 CAGTCGGTCTACGTTGCTACTGATTCCGAGAGTTATGTGCCTGAGCTCCAACAGCTCTTC 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 CAGTCGGTCTACGTTGCTACTGATTCCGAGAGTTATGTGCCTGAGCTCCAACAGCTCTTC 900

Qy        901 AAAGGGAAGGTGAAGGTGGTGAGCCTGAAGCCTGAGGTGGCCCAGGTCGACCTGTACATC 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 AAAGGGAAGGTGAAGGTGGTGAGCCTGAAGCCTGAGGTGGCCCAGGTCGACCTGTACATC 960

Qy        961 CTCGGCCAAGCCGACCACTTTATTGGCAACTGTGTCTCCTCCTTCACTGCCTTTGTGAAG 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 CTCGGCCAAGCCGACCACTTTATTGGCAACTGTGTCTCCTCCTTCACTGCCTTTGTGAAG 1020

Qy       1021 CGGGAGCGGGACCTCCAGGGGAGGCCGTCTTCTTTCTTCGGCATGGACAGGCCCCCTAAG 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1021 CGGGAGCGGGACCTCCAGGGGAGGCCGTCTTCTTTCTTCGGCATGGACAGGCCCCCTAAG 1080

Qy       1081 CTGCGGGACGAGTTCTGATT 1100
              ||||||||||||||||||||
Db       1081 CTGCGGGACGAGTTCTGATT 1100


RESULT 2
BD103196
LOCUS       BD103196                1100 bp    DNA     linear   PAT 27-AUG-2002
DEFINITION  O-fucosyltransferase.
ACCESSION   BD103196
VERSION     BD103196.1  GI:22648770
KEYWORDS    JP 2001527389-A/9.
SOURCE      unidentified
  ORGANISM  unidentified
            unclassified.
REFERENCE   1  (bases 1 to 1100)
  AUTHORS   Wang,Y. and Spellman,M.W.
  TITLE     O-fucosyltransferase
  JOURNAL   Patent: JP 2001527389-A 9 25-DEC-2001;
            GENENTECH INC
COMMENT     OS   Unidentified
            PN   JP 2001527389-A/9
            PD   25-DEC-2001
            PF   17-DEC-1997 JP 1998532877
            PR   31-JAN-1997 US    08/792498,26-NOV-1997 US    08/978741 PI
            YANG WANG,MICHAEL W SPELLMAN
            PC   C12N15/54,C12N9/10,C07K16/40
            CC   Strandedness: Single;
            CC   Topology: Linear;
            CC   O-fucosyltransferase
            FH   Key             Location/Qualifiers
            FT   source          1. .1100
            FT                   /organism='Unidentified'.
FEATURES             Location/Qualifiers
     source          1. .1100
                     /organism="unidentified"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:32644"
ORIGIN

  Query Match           100.0%;  Score 1100;  DB 6;  Length 1100;
  Best Local Similarity 100.0%;  Pred. No. 1e-253;
  Matches 1100;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 ATGCCCGCGGGCTCCTGGGACCCGGCCGGTTACCTGCTCTACTGCCCCTGCATGGGGCGC 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGCCCGCGGGCTCCTGGGACCCGGCCGGTTACCTGCTCTACTGCCCCTGCATGGGGCGC 60

Qy         61 TTTGGGAACCAGGCCGATCACTTCTTGGGCTCTCTGGCATTTGCAAAGCTGCTAAACCGT 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 TTTGGGAACCAGGCCGATCACTTCTTGGGCTCTCTGGCATTTGCAAAGCTGCTAAACCGT 120

Qy        121 ACCTTGGCTGTCCCTCCTTGGATTGAGTACCAGCATCACAAGCCTCCTTTCACCAACCTC 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 ACCTTGGCTGTCCCTCCTTGGATTGAGTACCAGCATCACAAGCCTCCTTTCACCAACCTC 180

Qy        181 CATGTGTCCTACCAGAAGTACTTCAAGCTGGAGCCCCTCCAGGCTTACCATCGGGTCATC 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 CATGTGTCCTACCAGAAGTACTTCAAGCTGGAGCCCCTCCAGGCTTACCATCGGGTCATC 240

Qy        241 AGCTTGGAGGATTTCATGGAGAAGCTGGCACCCACCCACTGGCCCCCTGAGAAGCGGGTG 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 AGCTTGGAGGATTTCATGGAGAAGCTGGCACCCACCCACTGGCCCCCTGAGAAGCGGGTG 300

Qy        301 GCATACTGCTTTGAGGTGGCAGCCCAGCGAAGCCCAGATAAGAAGACGTGCCCCATGAAG 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 GCATACTGCTTTGAGGTGGCAGCCCAGCGAAGCCCAGATAAGAAGACGTGCCCCATGAAG 360

Qy        361 GAAGGAAACCCCTTTGGCCCATTCTGGGATCAGTTTCATGTGAGTTTCAACAAGTCGGAG 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 GAAGGAAACCCCTTTGGCCCATTCTGGGATCAGTTTCATGTGAGTTTCAACAAGTCGGAG 420

Qy        421 CTTTTTACAGGCATTTCCTTCAGTGCTTCCTACAGAGAACAATGGAGCCAGAGATTTTCT 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 CTTTTTACAGGCATTTCCTTCAGTGCTTCCTACAGAGAACAATGGAGCCAGAGATTTTCT 480

Qy        481 CCAAAGGAACATCCGGTGCTTGCCCTGCCAGGAGCCCCAGCCCAGTTCCCCGTCCTAGAA 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 CCAAAGGAACATCCGGTGCTTGCCCTGCCAGGAGCCCCAGCCCAGTTCCCCGTCCTAGAA 540

Qy        541 GAACACAGGCCACTACAGAAGTACATGGTATGGTCAGACGAAATGGTGAAGACGGGAGAG 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 GAACACAGGCCACTACAGAAGTACATGGTATGGTCAGACGAAATGGTGAAGACGGGAGAG 600

Qy        601 GCCCAGATTCATGCCCACCTTGTCCGGCCCTATGTGGGCATTCATCTGCGCATTGGCTCT 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 GCCCAGATTCATGCCCACCTTGTCCGGCCCTATGTGGGCATTCATCTGCGCATTGGCTCT 660

Qy        661 GACTGGAAGAACGCCTGTGCCATGCTGAAGGACGGGACTGCAGGCTCGCACTTCATGGCC 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 GACTGGAAGAACGCCTGTGCCATGCTGAAGGACGGGACTGCAGGCTCGCACTTCATGGCC 720
```

```
Qy        721 TCTCCGCAGTGTGTGGGCTACAGCCGCAGCACAGCGGCCCCCCTCACGATGACTATGTGC 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 TCTCCGCAGTGTGTGGGCTACAGCCGCAGCACAGCGGCCCCCCTCACGATGACTATGTGC 780

Qy        781 CTGCCTGACCTGAAGGAGATCCAGAGGGCTGTGAAGCTCTGGGTGAGGTCGCTGGATGCC 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 CTGCCTGACCTGAAGGAGATCCAGAGGGCTGTGAAGCTCTGGGTGAGGTCGCTGGATGCC 840

Qy        841 CAGTCGGTCTACGTTGCTACTGATTCCGAGAGTTATGTGCCTGAGCTCCAACAGCTCTTC 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 CAGTCGGTCTACGTTGCTACTGATTCCGAGAGTTATGTGCCTGAGCTCCAACAGCTCTTC 900

Qy        901 AAAGGGAAGGTGAAGGTGGTGAGCCTGAAGCCTGAGGTGGCCCAGGTCGACCTGTACATC 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 AAAGGGAAGGTGAAGGTGGTGAGCCTGAAGCCTGAGGTGGCCCAGGTCGACCTGTACATC 960

Qy        961 CTCGGCCAAGCCGACCACTTTATTGGCAACTGTGTCTCCTCCTTCACTGCCTTTGTGAAG 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 CTCGGCCAAGCCGACCACTTTATTGGCAACTGTGTCTCCTCCTTCACTGCCTTTGTGAAG 1020

Qy       1021 CGGGAGCGGGACCTCCAGGGGAGGCCGTCTTCTTTCTTCGGCATGGACAGGCCCCCTAAG 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1021 CGGGAGCGGGACCTCCAGGGGAGGCCGTCTTCTTTCTTCGGCATGGACAGGCCCCCTAAG 1080

Qy       1081 CTGCGGGACGAGTTCTGATT 1100
              ||||||||||||||||||||
Db       1081 CTGCGGGACGAGTTCTGATT 1100
```

```
RESULT 3
AR163451
LOCUS       AR163451               1514 bp    DNA     linear   PAT 17-OCT-2001
DEFINITION  Sequence 2 from patent US 6270987.
ACCESSION   AR163451
VERSION     AR163451.1  GI:16234056
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 1514)
  AUTHORS   Wang,Y. and Spellman,M.W.
  TITLE     O-fucosyltransferase
  JOURNAL   Patent: US 6270987-A 2 07-AUG-2001;
FEATURES             Location/Qualifiers
     source          1. .1514
                     /organism="unknown"
                     /mol_type="unassigned DNA"
ORIGIN

  Query Match             100.0%;  Score 1100;  DB 6;  Length 1514;
  Best Local Similarity   100.0%;  Pred. No. 1e-253;
  Matches 1100;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 ATGCCCGCGGGCTCCTGGGACCCGGCCGGTTACCTGCTCTACTGCCCCTGCATGGGGCGC 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGCCCGCGGGCTCCTGGGACCCGGCCGGTTACCTGCTCTACTGCCCCTGCATGGGGCGC 60

Qy         61 TTTGGGAACCAGGCCGATCACTTCTTGGGCTCTCTGGCATTTGCAAAGCTGCTAAACCGT 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 TTTGGGAACCAGGCCGATCACTTCTTGGGCTCTCTGGCATTTGCAAAGCTGCTAAACCGT 120

Qy        121 ACCTTGGCTGTCCCTCCTTGGATTGAGTACCAGCATCACAAGCCTCCTTTCACCAACCTC 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 ACCTTGGCTGTCCCTCCTTGGATTGAGTACCAGCATCACAAGCCTCCTTTCACCAACCTC 180

Qy        181 CATGTGTCCTACCAGAAGTACTTCAAGCTGGAGCCCCTCCAGGCTTACCATCGGGTCATC 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 CATGTGTCCTACCAGAAGTACTTCAAGCTGGAGCCCCTCCAGGCTTACCATCGGGTCATC 240

Qy        241 AGCTTGGAGGATTTCATGGAGAAGCTGGCACCCACCCACTGGCCCCCTGAGAAGCGGGTG 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 AGCTTGGAGGATTTCATGGAGAAGCTGGCACCCACCCACTGGCCCCCTGAGAAGCGGGTG 300

Qy        301 GCATACTGCTTTGAGGTGGCAGCCCAGCGAAGCCCAGATAAGAAGACGTGCCCCATGAAG 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 GCATACTGCTTTGAGGTGGCAGCCCAGCGAAGCCCAGATAAGAAGACGTGCCCCATGAAG 360

Qy        361 GAAGGAAACCCCTTTGGCCCATTCTGGGATCAGTTTCATGTGAGTTTCAACAAGTCGGAG 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 GAAGGAAACCCCTTTGGCCCATTCTGGGATCAGTTTCATGTGAGTTTCAACAAGTCGGAG 420

Qy        421 CTTTTTACAGGCATTTCCTTCAGTGCTTCCTACAGAGAACAATGGAGCCAGAGATTTTCT 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 CTTTTTACAGGCATTTCCTTCAGTGCTTCCTACAGAGAACAATGGAGCCAGAGATTTTCT 480

Qy        481 CCAAAGGAACATCCGGTGCTTGCCCTGCCAGGAGCCCCAGCCCAGTTCCCCGTCCTAGAA 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 CCAAAGGAACATCCGGTGCTTGCCCTGCCAGGAGCCCCAGCCCAGTTCCCCGTCCTAGAA 540

Qy        541 GAACACAGGCCACTACAGAAGTACATGGTATGGTCAGACGAAATGGTGAAGACGGGAGAG 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 GAACACAGGCCACTACAGAAGTACATGGTATGGTCAGACGAAATGGTGAAGACGGGAGAG 600

Qy        601 GCCCAGATTCATGCCCACCTTGTCCGGCCCTATGTGGGCATTCATCTGCGCATTGGCTCT 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 GCCCAGATTCATGCCCACCTTGTCCGGCCCTATGTGGGCATTCATCTGCGCATTGGCTCT 660

Qy        661 GACTGGAAGAACGCCTGTGCCATGCTGAAGGACGGGACTGCAGGCTCGCACTTCATGGCC 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 GACTGGAAGAACGCCTGTGCCATGCTGAAGGACGGGACTGCAGGCTCGCACTTCATGGCC 720

Qy        721 TCTCCGCAGTGTGTGGGCTACAGCCGCAGCACAGCGGCCCCCCTCACGATGACTATGTGC 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 TCTCCGCAGTGTGTGGGCTACAGCCGCAGCACAGCGGCCCCCCTCACGATGACTATGTGC 780

Qy        781 CTGCCTGACCTGAAGGAGATCCAGAGGGCTGTGAAGCTCTGGGTGAGGTCGCTGGATGCC 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 CTGCCTGACCTGAAGGAGATCCAGAGGGCTGTGAAGCTCTGGGTGAGGTCGCTGGATGCC 840

Qy        841 CAGTCGGTCTACGTTGCTACTGATTCCGAGAGTTATGTGCCTGAGCTCCAACAGCTCTTC 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 CAGTCGGTCTACGTTGCTACTGATTCCGAGAGTTATGTGCCTGAGCTCCAACAGCTCTTC 900

Qy        901 AAAGGGAAGGTGAAGGTGGTGAGCCTGAAGCCTGAGGTGGCCCAGGTCGACCTGTACATC 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 AAAGGGAAGGTGAAGGTGGTGAGCCTGAAGCCTGAGGTGGCCCAGGTCGACCTGTACATC 960

Qy        961 CTCGGCCAAGCCGACCACTTTATTGGCAACTGTGTCTCCTCCTTCACTGCCTTTGTGAAG 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 CTCGGCCAAGCCGACCACTTTATTGGCAACTGTGTCTCCTCCTTCACTGCCTTTGTGAAG 1020

Qy       1021 CGGGAGCGGGACCTCCAGGGGAGGCCGTCTTCTTTCTTCGGCATGGACAGGCCCCCTAAG 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1021 CGGGAGCGGGACCTCCAGGGGAGGCCGTCTTCTTTCTTCGGCATGGACAGGCCCCCTAAG 1080

Qy       1081 CTGCGGGACGAGTTCTGATT 1100
              ||||||||||||||||||||
Db       1081 CTGCGGGACGAGTTCTGATT 1100
```

```
RESULT 4
BD103188
LOCUS       BD103188               1514 bp    DNA     linear   PAT 27-AUG-2002
DEFINITION  O-fucosyltransferase.
ACCESSION   BD103188
VERSION     BD103188.1  GI:22648762
KEYWORDS    JP 2001527389-A/1.
SOURCE      unidentified
  ORGANISM  unidentified
            unclassified.
REFERENCE   1  (bases 1 to 1514)
  AUTHORS   Wang,Y. and Spellman,M.W.
  TITLE     O-fucosyltransferase
  JOURNAL   Patent: JP 2001527389-A 1 25-DEC-2001;
            GENENTECH INC
COMMENT     OS   Unidentified
```

```
            PN   JP 2001527389-A/1
            PD   25-DEC-2001
            PF   17-DEC-1997 JP 1998532877
            PR   31-JAN-1997 US    08/792498,26-NOV-1997 US    08/978741 PI
            YANG WANG,MICHAEL W SPELLMAN
            PC   C12N15/54,C12N9/10,C07K16/40
            CC   Strandedness: Single;
            CC   Topology: Linear;
            CC   O-fucosyltransferase
            FH   Key             Location/Qualifiers
            FT   source          1. .1514
            FT                   /organism='Unidentified'.
FEATURES             Location/Qualifiers
     source          1. .1514
                     /organism="unidentified"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:32644"
ORIGIN

  Query Match             100.0%;  Score 1100;  DB 6;  Length 1514;
  Best Local Similarity   100.0%;  Pred. No. 1e-253;
  Matches 1100;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 ATGCCCGCGGGCTCCTGGGACCCGGCCGGTTACCTGCTCTACTGCCCCTGCATGGGGCGC 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGCCCGCGGGCTCCTGGGACCCGGCCGGTTACCTGCTCTACTGCCCCTGCATGGGGCGC 60

Qy         61 TTTGGGAACCAGGCCGATCACTTCTTGGGCTCTCTGGCATTTGCAAAGCTGCTAAACCGT 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 TTTGGGAACCAGGCCGATCACTTCTTGGGCTCTCTGGCATTTGCAAAGCTGCTAAACCGT 120

Qy        121 ACCTTGGCTGTCCCTCCTTGGATTGAGTACCAGCATCACAAGCCTCCTTTCACCAACCTC 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 ACCTTGGCTGTCCCTCCTTGGATTGAGTACCAGCATCACAAGCCTCCTTTCACCAACCTC 180

Qy        181 CATGTGTCCTACCAGAAGTACTTCAAGCTGGAGCCCCTCCAGGCTTACCATCGGGTCATC 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 CATGTGTCCTACCAGAAGTACTTCAAGCTGGAGCCCCTCCAGGCTTACCATCGGGTCATC 240

Qy        241 AGCTTGGAGGATTTCATGGAGAAGCTGGCACCCACCCACTGGCCCCCTGAGAAGCGGGTG 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 AGCTTGGAGGATTTCATGGAGAAGCTGGCACCCACCCACTGGCCCCCTGAGAAGCGGGTG 300

Qy        301 GCATACTGCTTTGAGGTGGCAGCCCAGCGAAGCCCAGATAAGAAGACGTGCCCCATGAAG 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 GCATACTGCTTTGAGGTGGCAGCCCAGCGAAGCCCAGATAAGAAGACGTGCCCCATGAAG 360

Qy        361 GAAGGAAACCCCTTTGGCCCATTCTGGGATCAGTTTCATGTGAGTTTCAACAAGTCGGAG 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 GAAGGAAACCCCTTTGGCCCATTCTGGGATCAGTTTCATGTGAGTTTCAACAAGTCGGAG 420

Qy        421 CTTTTTACAGGCATTTCCTTCAGTGCTTCCTACAGAGAACAATGGAGCCAGAGATTTTCT 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 CTTTTTACAGGCATTTCCTTCAGTGCTTCCTACAGAGAACAATGGAGCCAGAGATTTTCT 480

Qy        481 CCAAAGGAACATCCGGTGCTTGCCCTGCCAGGAGCCCCAGCCCAGTTCCCCGTCCTAGAA 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 CCAAAGGAACATCCGGTGCTTGCCCTGCCAGGAGCCCCAGCCCAGTTCCCCGTCCTAGAA 540

Qy        541 GAACACAGGCCACTACAGAAGTACATGGTATGGTCAGACGAAATGGTGAAGACGGGAGAG 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 GAACACAGGCCACTACAGAAGTACATGGTATGGTCAGACGAAATGGTGAAGACGGGAGAG 600

Qy        601 GCCCAGATTCATGCCCACCTTGTCCGGCCCTATGTGGGCATTCATCTGCGCATTGGCTCT 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 GCCCAGATTCATGCCCACCTTGTCCGGCCCTATGTGGGCATTCATCTGCGCATTGGCTCT 660

Qy        661 GACTGGAAGAACGCCTGTGCCATGCTGAAGGACGGGACTGCAGGCTCGCACTTCATGGCC 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 GACTGGAAGAACGCCTGTGCCATGCTGAAGGACGGGACTGCAGGCTCGCACTTCATGGCC 720

Qy        721 TCTCCGCAGTGTGTGGGCTACAGCCGCAGCACAGCGGCCCCCCTCACGATGACTATGTGC 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 TCTCCGCAGTGTGTGGGCTACAGCCGCAGCACAGCGGCCCCCCTCACGATGACTATGTGC 780

Qy        781 CTGCCTGACCTGAAGGAGATCCAGAGGGCTGTGAAGCTCTGGGTGAGGTCGCTGGATGCC 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 CTGCCTGACCTGAAGGAGATCCAGAGGGCTGTGAAGCTCTGGGTGAGGTCGCTGGATGCC 840

Qy        841 CAGTCGGTCTACGTTGCTACTGATTCCGAGAGTTATGTGCCTGAGCTCCAACAGCTCTTC 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 CAGTCGGTCTACGTTGCTACTGATTCCGAGAGTTATGTGCCTGAGCTCCAACAGCTCTTC 900

Qy        901 AAAGGGAAGGTGAAGGTGGTGAGCCTGAAGCCTGAGGTGGCCCAGGTCGACCTGTACATC 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 AAAGGGAAGGTGAAGGTGGTGAGCCTGAAGCCTGAGGTGGCCCAGGTCGACCTGTACATC 960

Qy        961 CTCGGCCAAGCCGACCACTTTATTGGCAACTGTGTCTCCTCCTTCACTGCCTTTGTGAAG 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 CTCGGCCAAGCCGACCACTTTATTGGCAACTGTGTCTCCTCCTTCACTGCCTTTGTGAAG 1020

Qy       1021 CGGGAGCGGGACCTCCAGGGGAGGCCGTCTTCTTTCTTCGGCATGGACAGGCCCCCTAAG 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1021 CGGGAGCGGGACCTCCAGGGGAGGCCGTCTTCTTTCTTCGGCATGGACAGGCCCCCTAAG 1080

Qy       1081 CTGCGGGACGAGTTCTGATT 1100
              ||||||||||||||||||||
Db       1081 CTGCGGGACGAGTTCTGATT 1100


RESULT 5
AR163452
LOCUS       AR163452                1300 bp    DNA     linear   PAT 17-OCT-2001
DEFINITION  Sequence 6 from patent US 6270987.
ACCESSION   AR163452
VERSION     AR163452.1  GI:16234057
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 1300)
  AUTHORS   Wang,Y. and Spellman,M.W.
  TITLE     O-fucosyltransferase
  JOURNAL   Patent: US 6270987-A 6 07-AUG-2001;
FEATURES             Location/Qualifiers
     source          1. .1300
                     /organism="unknown"
                     /mol_type="unassigned DNA"
ORIGIN

  Query Match             99.9%;  Score 1098.4;  DB 6;  Length 1300;
  Best Local Similarity   99.9%;  Pred. No. 2.4e-253;
  Matches 1099;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 ATGCCCGCGGGCTCCTGGGACCCGGCCGGTTACCTGCTCTACTGCCCCTGCATGGGGCGC 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        136 ATGCCCGCGGGCTCCTGGGACCCGGCCGGTTACCTGCTCTACTGCCCCTGCATGGGGCGC 195

Qy         61 TTTGGGAACCAGGCCGATCACTTCTTGGGCTCTCTGGCATTTGCAAAGCTGCTAAACCGT 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        196 TTTGGGAACCAGGCCGATCACTTCTTGGGCTCTCTGGCATTTGCAAAGCTGCTAAACCGT 255

Qy        121 ACCTTGGCTGTCCCTCCTTGGATTGAGTACCAGCATCACAAGCCTCCTTTCACCAACCTC 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        256 ACCTTGGCTGTCCCTCCTTGGATTGAGTACCAGCATCACAAGCCTCCTTTCACCAACCTC 315

Qy        181 CATGTGTCCTACCAGAAGTACTTCAAGCTGGAGCCCCTCCAGGCTTACCATCGGGTCATC 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        316 CATGTGTCCTACCAGAAGTACTTCAAGCTGGAGCCCCTCCAGGCTTACCATCGGGTCATC 375

Qy        241 AGCTTGGAGGATTTCATGGAGAAGCTGGCACCCACCCACTGGCCCCCTGAGAAGCGGGTG 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        376 AGCTTGGAGGATTTCATGGAGAAGCTGGCACCCACCCACTGGCCCCCTGAGAAGCGGGTG 435

Qy        301 GCATACTGCTTTGAGGTGGCAGCCCAGCGAAGCCCAGATAAGAAGACGTGCCCCATGAAG 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db        436 GCATACTGCTTTGAGGTGGCAGCCCAGCGAAGCCCAGATAAGAAGACGTGCCCCATGAAG 495

Qy        361 GAAGGAAACCCCTTTGGCCCATTCTGGGATCAGTTTCATGTGAGTTTCAACAAGTCGGAG 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        496 GAAGGAAACCCCTTTGGCCCATTCTGGGATCAGTTTCATGTGAGTTTCAACAAGTCGGAG 555

Qy        421 CTTTTTACAGGCATTTCCTTCAGTGCTTCCTACAGAGAACAATGGAGCCAGAGATTTTCT 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        556 CTTTTTACAGGCATTTCCTTCAGTGCTTCCTACAGAGAACAATGGAGCCAGAGATTTTCT 615

Qy        481 CCAAAGGAACATCCGGTGCTTGCCCTGCCAGGAGCCCCAGCCCAGTTCCCCGTCCTAGAA 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        616 CCAAAGGAACATCCGGTGCTTGCCCTGCCAGGAGCCCCAGCCCAGTTCCCCGTCCTAGAG 675

Qy        541 GAACACAGGCCACTACAGAAGTACATGGTATGGTCAGACGAAATGGTGAAGACGGGAGAG 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        676 GAACACAGGCCACTACAGAAGTACATGGTATGGTCAGACGAAATGGTGAAGACGGGAGAG 735

Qy        601 GCCCAGATTCATGCCCACCTTGTCCGGCCCTATGTGGGCATTCATCTGCGCATTGGCTCT 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        736 GCCCAGATTCATGCCCACCTTGTCCGGCCCTATGTGGGCATTCATCTGCGCATTGGCTCT 795

Qy        661 GACTGGAAGAACGCCTGTGCCATGCTGAAGGACGGGACTGCAGGCTCGCACTTCATGGCC 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        796 GACTGGAAGAACGCCTGTGCCATGCTGAAGGACGGGACTGCAGGCTCGCACTTCATGGCC 855

Qy        721 TCTCCGCAGTGTGTGGGCTACAGCCGCAGCACAGCGGCCCCCCTCACGATGACTATGTGC 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        856 TCTCCGCAGTGTGTGGGCTACAGCCGCAGCACAGCGGCCCCCCTCACGATGACTATGTGC 915

Qy        781 CTGCCTGACCTGAAGGAGATCCAGAGGGCTGTGAAGCTCTGGGTGAGGTCGCTGGATGCC 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        916 CTGCCTGACCTGAAGGAGATCCAGAGGGCTGTGAAGCTCTGGGTGAGGTCGCTGGATGCC 975

Qy        841 CAGTCGGTCTACGTTGCTACTGATTCCGAGAGTTATGTGCCTGAGCTCCAACAGCTCTTC 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        976 CAGTCGGTCTACGTTGCTACTGATTCCGAGAGTTATGTGCCTGAGCTCCAACAGCTCTTC 1035

Qy        901 AAAGGGAAGGTGAAGGTGGTGAGCCTGAAGCCTGAGGTGGCCCAGGTCGACCTGTACATC 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1036 AAAGGGAAGGTGAAGGTGGTGAGCCTGAAGCCTGAGGTGGCCCAGGTCGACCTGTACATC 1095

Qy        961 CTCGGCCAAGCCGACCACTTTATTGGCAACTGTGTCTCCTCCTTCACTGCCTTTGTGAAG 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1096 CTCGGCCAAGCCGACCACTTTATTGGCAACTGTGTCTCCTCCTTCACTGCCTTTGTGAAG 1155

Qy       1021 CGGGAGCGGGACCTCCAGGGGAGGCCGTCTTCTTTCTTCGGCATGGACAGGCCCCCTAAG 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1156 CGGGAGCGGGACCTCCAGGGGAGGCCGTCTTCTTTCTTCGGCATGGACAGGCCCCCTAAG 1215

Qy       1081 CTGCGGGACGAGTTCTGATT 1100
              ||||||||||||||||||||
Db       1216 CTGCGGGACGAGTTCTGATT 1235


RESULT 6
BD103189
LOCUS       BD103189               1300 bp    DNA     linear   PAT 27-AUG-2002
DEFINITION  O-fucosyltransferase.
ACCESSION   BD103189
VERSION     BD103189.1  GI:22648763
KEYWORDS    JP 2001527389-A/2.
SOURCE      unidentified
  ORGANISM  unidentified
            unclassified.
REFERENCE   1  (bases 1 to 1300)
  AUTHORS   Wang,Y. and Spellman,M.W.
  TITLE     O-fucosyltransferase
  JOURNAL   Patent: JP 2001527389-A 2 25-DEC-2001;
            GENENTECH INC
COMMENT     OS   Unidentified
            PN   JP 2001527389-A/2
            PD   25-DEC-2001
            PF   17-DEC-1997 JP 1998532877
            PR   31-JAN-1997 US    08/792498,26-NOV-1997 US    08/978741 PI
            YANG WANG,MICHAEL W SPELLMAN
            PC   C12N15/54,C12N9/10,C07K16/40
            CC   Strandedness: Single;
            CC   Topology: Linear;
            CC   O-fucosyltransferase
            FH   Key             Location/Qualifiers
            FT   source          1. .1300
            FT                   /organism='Unidentified'.
FEATURES             Location/Qualifiers
     source          1. .1300
                     /organism="unidentified"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:32644"
ORIGIN

  Query Match             99.9%;  Score 1098.4;  DB 6;  Length 1300;
  Best Local Similarity   99.9%;  Pred. No. 2.4e-253;
  Matches 1099;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 ATGCCCGCGGGCTCCTGGGACCCGGCCGGTTACCTGCTCTACTGCCCCTGCATGGGGCGC 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        136 ATGCCCGCGGGCTCCTGGGACCCGGCCGGTTACCTGCTCTACTGCCCCTGCATGGGGCGC 195

Qy         61 TTTGGGAACCAGGCCGATCACTTCTTGGGCTCTCTGGCATTTGCAAAGCTGCTAAACCGT 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        196 TTTGGGAACCAGGCCGATCACTTCTTGGGCTCTCTGGCATTTGCAAAGCTGCTAAACCGT 255

Qy        121 ACCTTGGCTGTCCCTCCTTGGATTGAGTACCAGCATCACAAGCCTCCTTTCACCAACCTC 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        256 ACCTTGGCTGTCCCTCCTTGGATTGAGTACCAGCATCACAAGCCTCCTTTCACCAACCTC 315

Qy        181 CATGTGTCCTACCAGAAGTACTTCAAGCTGGAGCCCCTCCAGGCTTACCATCGGGTCATC 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        316 CATGTGTCCTACCAGAAGTACTTCAAGCTGGAGCCCCTCCAGGCTTACCATCGGGTCATC 375

Qy        241 AGCTTGGAGGATTTCATGGAGAAGCTGGCACCCACCCACTGGCCCCCTGAGAAGCGGGTG 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        376 AGCTTGGAGGATTTCATGGAGAAGCTGGCACCCACCCACTGGCCCCCTGAGAAGCGGGTG 435

Qy        301 GCATACTGCTTTGAGGTGGCAGCCCAGCGAAGCCCAGATAAGAAGACGTGCCCCATGAAG 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        436 GCATACTGCTTTGAGGTGGCAGCCCAGCGAAGCCCAGATAAGAAGACGTGCCCCATGAAG 495

Qy        361 GAAGGAAACCCCTTTGGCCCATTCTGGGATCAGTTTCATGTGAGTTTCAACAAGTCGGAG 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        496 GAAGGAAACCCCTTTGGCCCATTCTGGGATCAGTTTCATGTGAGTTTCAACAAGTCGGAG 555

Qy        421 CTTTTTACAGGCATTTCCTTCAGTGCTTCCTACAGAGAACAATGGAGCCAGAGATTTTCT 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        556 CTTTTTACAGGCATTTCCTTCAGTGCTTCCTACAGAGAACAATGGAGCCAGAGATTTTCT 615

Qy        481 CCAAAGGAACATCCGGTGCTTGCCCTGCCAGGAGCCCCAGCCCAGTTCCCCGTCCTAGAA 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        616 CCAAAGGAACATCCGGTGCTTGCCCTGCCAGGAGCCCCAGCCCAGTTCCCCGTCCTAGAG 675

Qy        541 GAACACAGGCCACTACAGAAGTACATGGTATGGTCAGACGAAATGGTGAAGACGGGAGAG 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        676 GAACACAGGCCACTACAGAAGTACATGGTATGGTCAGACGAAATGGTGAAGACGGGAGAG 735

Qy        601 GCCCAGATTCATGCCCACCTTGTCCGGCCCTATGTGGGCATTCATCTGCGCATTGGCTCT 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        736 GCCCAGATTCATGCCCACCTTGTCCGGCCCTATGTGGGCATTCATCTGCGCATTGGCTCT 795

Qy        661 GACTGGAAGAACGCCTGTGCCATGCTGAAGGACGGGACTGCAGGCTCGCACTTCATGGCC 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        796 GACTGGAAGAACGCCTGTGCCATGCTGAAGGACGGGACTGCAGGCTCGCACTTCATGGCC 855

Qy        721 TCTCCGCAGTGTGTGGGCTACAGCCGCAGCACAGCGGCCCCCCTCACGATGACTATGTGC 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        856 TCTCCGCAGTGTGTGGGCTACAGCCGCAGCACAGCGGCCCCCCTCACGATGACTATGTGC 915
```

```
Qy        781 CTGCCTGACCTGAAGGAGATCCAGAGGGCTGTGAAGCTCTGGGTGAGGTCGCTGGATGCC 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        916 CTGCCTGACCTGAAGGAGATCCAGAGGGCTGTGAAGCTCTGGGTGAGGTCGCTGGATGCC 975

Qy        841 CAGTCGGTCTACGTTGCTACTGATTCCGAGAGTTATGTGCCTGAGCTCCAACAGCTCTTC 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        976 CAGTCGGTCTACGTTGCTACTGATTCCGAGAGTTATGTGCCTGAGCTCCAACAGCTCTTC 1035

Qy        901 AAAGGGAAGGTGAAGGTGGTGAGCCTGAAGCCTGAGGTGGCCCAGGTCGACCTGTACATC 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1036 AAAGGGAAGGTGAAGGTGGTGAGCCTGAAGCCTGAGGTGGCCCAGGTCGACCTGTACATC 1095

Qy        961 CTCGGCCAAGCCGACCACTTTATTGGCAACTGTGTCTCCTCCTTCACTGCCTTTGTGAAG 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1096 CTCGGCCAAGCCGACCACTTTATTGGCAACTGTGTCTCCTCCTTCACTGCCTTTGTGAAG 1155

Qy       1021 CGGGAGCGGGACCTCCAGGGGAGGCCGTCTTCTTTCTTCGGCATGGACAGGCCCCCTAAG 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1156 CGGGAGCGGGACCTCCAGGGGAGGCCGTCTTCTTTCTTCGGCATGGACAGGCCCCCTAAG 1215

Qy       1081 CTGCGGGACGAGTTCTGATT 1100
              ||||||||||||||||||||
Db       1216 CTGCGGGACGAGTTCTGATT 1235
```

```
RESULT 7
AF375884
LOCUS       AF375884                5249 bp    mRNA    linear   PRI 23-OCT-2001
DEFINITION  Homo sapiens protein o-fucosyltransferase (POFUT1) mRNA, complete
            cds.
ACCESSION   AF375884
VERSION     AF375884.1  GI:15825113
KEYWORDS    .
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 5249)
  AUTHORS   Wang,Y., Shao,L., Shi,S., Harris,R.J., Spellman,M.W., Stanley,P.
            and Haltiwanger,R.S.
  TITLE     Modification of epidermal growth factor-like repeats with O-fucose.
            Molecular cloning and expression of a novel GDP-fucose protein
            O-fucosyltransferase
  JOURNAL   J. Biol. Chem. 276 (43), 40338-40345 (2001)
  MEDLINE   21523965
   PUBMED   11524432
REFERENCE   2  (bases 1 to 5249)
  AUTHORS   Shi,S., Stanley,P., Wang,Y., Shao,L., Harris,R.J., Spellman,M.W.
            and Haltiwanger,R.S.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-MAY-2001) Cell Biology, Albert Einstein College of
            Medicine, 1300 Morris Park Ave., New York, NY 10461, USA
FEATURES             Location/Qualifiers
     source          1. .5249
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /chromosome="20"
                     /map="between PLAGL2 and KIF3B"
                     /tissue_type="heart"
     gene            1. .5249
                     /gene="POFUT1"
     CDS             50. .1216
                     /gene="POFUT1"
                     /function="catalyzes transfer of fucose from GDP-fucose to
                     serine or threonine on epidermal growth factor-like
                     repeats"
                     /note="O-FucT-1; glycosyltransferase"
                     /codon_start=1
                     /product="protein o-fucosyltransferase"
                     /protein_id="AAL09576.1"
                     /db_xref="GI:15825114"
                     /translation="MGAAAWARPLSVSFLLLLLPLPGMPAGSWDPAGYLLYCPCMGRF
                     GNQADHFLGSLAFAKLLNRTLAVPPWIEYQHHKPPFTNLHVSYQKYFKLEPLQAYHRV
                     ISLEDFMEKLAPTHWPPEKRVAYCFEVAAQRSPDKKTCPMKEGNPFGPFWDQFHVSFN
                     KSELFTGISFSASYREQWSQRFSPKEHPVLALPGAPAQFPVLEEHRPLQKYMVWSDEM
                     VKTGEAQIHAHLVRPYVGIHLRIGSDWKNACAMLKDGTAGSHFMASPQCVGYSRSTAA
                     PLTMTMCLPDLKEIQRAVKLWVRSLDAQSVYVATDSESYVPELQQLFKGKVKVVSLKP
                     EVAQVDLYILGQADHFIGNCVSSFTAFVKRERDLQGRPSSFFGMDRPPKLRDEF"
ORIGIN

  Query Match             99.9%;  Score 1098.4;  DB 9;  Length 5249;
  Best Local Similarity   99.9%;  Pred. No. 2.4e-253;
  Matches 1099;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 ATGCCCGCGGGCTCCTGGGACCCGGCCGGTTACCTGCTCTACTGCCCCTGCATGGGGCGC 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        119 ATGCCCGCGGGCTCCTGGGACCCGGCCGGTTACCTGCTCTACTGCCCCTGCATGGGGCGC 178

Qy         61 TTTGGGAACCAGGCCGATCACTTCTTGGGCTCTCTGGCATTTGCAAAGCTGCTAAACCGT 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        179 TTTGGGAACCAGGCCGATCACTTCTTGGGCTCTCTGGCATTTGCAAAGCTGCTAAACCGT 238

Qy        121 ACCTTGGCTGTCCCTCCTTGGATTGAGTACCAGCATCACAAGCCTCCTTTCACCAACCTC 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        239 ACCTTGGCTGTCCCTCCTTGGATTGAGTACCAGCATCACAAGCCTCCTTTCACCAACCTC 298

Qy        181 CATGTGTCCTACCAGAAGTACTTCAAGCTGGAGCCCCTCCAGGCTTACCATCGGGTCATC 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        299 CATGTGTCCTACCAGAAGTACTTCAAGCTGGAGCCCCTCCAGGCTTACCATCGGGTCATC 358

Qy        241 AGCTTGGAGGATTTCATGGAGAAGCTGGCACCCACCCACTGGCCCCCTGAGAAGCGGGTG 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        359 AGCTTGGAGGATTTCATGGAGAAGCTGGCACCCACCCACTGGCCCCCTGAGAAGCGGGTG 418

Qy        301 GCATACTGCTTTGAGGTGGCAGCCCAGCGAAGCCCAGATAAGAAGACGTGCCCCATGAAG 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        419 GCATACTGCTTTGAGGTGGCAGCCCAGCGAAGCCCAGATAAGAAGACGTGCCCCATGAAG 478

Qy        361 GAAGGAAACCCCCTTTGGCCCATTCTGGGATCAGTTTCATGTGAGTTTCAACAAGTCGGAG 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        479 GAAGGAAACCCCCTTTGGCCCATTCTGGGATCAGTTTCATGTGAGTTTCAACAAGTCGGAG 538

Qy        421 CTTTTTACAGGCATTTCCTTCAGTGCTTCCTACAGAGAACAATGGAGCCAGAGATTTTCT 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        539 CTTTTTACAGGCATTTCCTTCAGTGCTTCCTACAGAGAACAATGGAGCCAGAGATTTTCT 598

Qy        481 CCAAAGGAACATCCGGTGCTTGCCCTGCCAGGAGCCCCAGCCCAGTTCCCCGTCCTAGAA 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        599 CCAAAGGAACATCCGGTGCTTGCCCTGCCAGGAGCCCCAGCCCAGTTCCCCGTCCTAGAG 658

Qy        541 GAACACAGGCCACTACAGAAGTACATGGTATGGTCAGACGAAATGGTGAAGACGGGAGAG 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        659 GAACACAGGCCACTACAGAAGTACATGGTATGGTCAGACGAAATGGTGAAGACGGGAGAG 718

Qy        601 GCCCAGATTCATGCCCACCTTGTCCGGCCCTATGTGGGCATTCATCTGCGCATTGGCTCT 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        719 GCCCAGATTCATGCCCACCTTGTCCGGCCCTATGTGGGCATTCATCTGCGCATTGGCTCT 778

Qy        661 GACTGGAAGAACGCCTGTGCCATGCTGAAGGACGGGACTGCAGGCTCGCACTTCATGGCC 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        779 GACTGGAAGAACGCCTGTGCCATGCTGAAGGACGGGACTGCAGGCTCGCACTTCATGGCC 838

Qy        721 TCTCCGCAGTGTGTGGGCTACAGCCGCAGCACAGCGGCCCCCCTCACGATGACTATGTGC 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        839 TCTCCGCAGTGTGTGGGCTACAGCCGCAGCACAGCGGCCCCCCTCACGATGACTATGTGC 898

Qy        781 CTGCCTGACCTGAAGGAGATCCAGAGGGCTGTGAAGCTCTGGGTGAGGTCGCTGGATGCC 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        899 CTGCCTGACCTGAAGGAGATCCAGAGGGCTGTGAAGCTCTGGGTGAGGTCGCTGGATGCC 958

Qy        841 CAGTCGGTCTACGTTGCTACTGATTCCGAGAGTTATGTGCCTGAGCTCCAACAGCTCTTC 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        959 CAGTCGGTCTACGTTGCTACTGATTCCGAGAGTTATGTGCCTGAGCTCCAACAGCTCTTC 1018

Qy        901 AAAGGGAAGGTGAAGGTGGTGAGCCTGAAGCCTGAGGTGGCCCAGGTCGACCTGTACATC 960
```

```
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1019 AAAGGGAAGGTGAAGGTGGTGAGCCTGAAGCCTGAGGTGGCCCAGGTCGACCTGTACATC 1078

Qy        961 CTCGGCCAAGCCGACCACTTTATTGGCAACTGTGTCTCCTCCTTCACTGCCTTTGTGAAG 1020
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1079 CTCGGCCAAGCCGACCACTTTATTGGCAACTGTGTCTCCTCCTTCACTGCCTTTGTGAAG 1138

Qy       1021 CGGGAGCGGGACCTCCAGGGGAGGCCGTCTTCTTTCTTCGGCATGGACAGGCCCCCTAAG 1080
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1139 CGGGAGCGGGACCTCCAGGGGAGGCCGTCTTCTTTCTTCGGCATGGACAGGCCCCCTAAG 1198

Qy       1081 CTGCGGGACGAGTTCTGATT 1100
                ||||||||||||||||||||
Db       1199 CTGCGGGACGAGTTCTGATT 1218


RESULT 8
BD103190
LOCUS       BD103190               11284 bp    DNA     linear   PAT 27-AUG-2002
DEFINITION  O-fucosyltransferase.
ACCESSION   BD103190
VERSION     BD103190.1  GI:22648764
KEYWORDS    JP 2001527389-A/3.
SOURCE      unidentified
  ORGANISM  unidentified
            unclassified.
REFERENCE   1  (bases 1 to 11284)
  AUTHORS   Wang,Y. and Spellman,M.W.
  TITLE     O-fucosyltransferase
  JOURNAL   Patent: JP 2001527389-A 3 25-DEC-2001;
            GENENTECH INC
COMMENT     OS   Unidentified
            PN   JP 2001527389-A/3
            PD   25-DEC-2001
            PF   17-DEC-1997 JP 1998532877
            PR   31-JAN-1997 US    08/792498,26-NOV-1997 US    08/978741 PI
            YANG WANG,MICHAEL W SPELLMAN
            PC   C12N15/54,C12N9/10,C07K16/40
            CC   Strandedness: Single;
            CC   Topology: Linear;
            CC   O-fucosyltransferase
            FH   Key             Location/Qualifiers
            FT   source          1. .11284
            FT                   /organism='Unidentified'.
FEATURES             Location/Qualifiers
     source          1. .11284
                     /organism="unidentified"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:32644"
ORIGIN

  Query Match             99.9%;  Score 1098.4;  DB 6;  Length 11284;
  Best Local Similarity   99.9%;  Pred. No. 2.4e-253;
  Matches 1099;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 ATGCCCGCGGGCTCCTGGGACCCGGCCGGTTACCTGCTCTACTGCCCCTGCATGGGGCGC 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4236 ATGCCCGCGGGCTCCTGGGACCCGGCCGGTTACCTGCTCTACTGCCCCTGCATGGGGCGC 4295

Qy         61 TTTGGGAACCAGGCCGATCACTTCTTGGGCTCTCTGGCATTTGCAAAGCTGCTAAACCGT 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4296 TTTGGGAACCAGGCCGATCACTTCTTGGGCTCTCTGGCATTTGCAAAGCTGCTAAACCGT 4355

Qy        121 ACCTTGGCTGTCCCTCCTTGGATTGAGTACCAGCATCACAAGCCTCCTTTCACCAACCTC 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4356 ACCTTGGCTGTCCCTCCTTGGATTGAGTACCAGCATCACAAGCCTCCTTTCACCAACCTC 4415

Qy        181 CATGTGTCCTACCAGAAGTACTTCAAGCTGGAGCCCCTCCAGGCTTACCATCGGGTCATC 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4416 CATGTGTCCTACCAGAAGTACTTCAAGCTGGAGCCCCTCCAGGCTTACCATCGGGTCATC 4475

Qy        241 AGCTTGGAGGATTTCATGGAGAAGCTGGCACCCACCCACTGGCCCCCTGAGAAGCGGGTG 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4476 AGCTTGGAGGATTTCATGGAGAAGCTGGCACCCACCCACTGGCCCCCTGAGAAGCGGGTG 4535

Qy        301 GCATACTGCTTTGAGGTGGCAGCCCAGCGAAGCCCAGATAAGAAGACGTGCCCCATGAAG 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4536 GCATACTGCTTTGAGGTGGCAGCCCAGCGAAGCCCAGATAAGAAGACGTGCCCCATGAAG 4595

Qy        361 GAAGGAAACCCCCTTTGGCCCATTCTGGGATCAGTTTCATGTGAGTTTCAACAAGTCGGAG 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4596 GAAGGAAACCCCCTTTGGCCCATTCTGGGATCAGTTTCATGTGAGTTTCAACAAGTCGGAG 4655

Qy        421 CTTTTTACAGGCATTTCCTTCAGTGCTTCCTACAGAGAACAATGGAGCCAGAGATTTTCT 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4656 CTTTTTACAGGCATTTCCTTCAGTGCTTCCTACAGAGAACAATGGAGCCAGAGATTTTCT 4715

Qy        481 CCAAAGGAACATCCGGTGCTTGCCCTGCCAGGAGCCCCAGCCCAGTTCCCCGTCCTAGAA 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4716 CCAAAGGAACATCCGGTGCTTGCCCTGCCAGGAGCCCCAGCCCAGTTCCCCGTCCTAGAG 4775

Qy        541 GAACACAGGCCACTACAGAAGTACATGGTATGGTCAGACGAAATGGTGAAGACGGGAGAG 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4776 GAACACAGGCCACTACAGAAGTACATGGTATGGTCAGACGAAATGGTGAAGACGGGAGAG 4835

Qy        601 GCCCAGATTCATGCCCACCTTGTCCGGCCCTATGTGGGCATTCATCTGCGCATTGGCTCT 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4836 GCCCAGATTCATGCCCACCTTGTCCGGCCCTATGTGGGCATTCATCTGCGCATTGGCTCT 4895

Qy        661 GACTGGAAGAACGCCTGTGCCATGCTGAAGGACGGGACTGCAGGCTCGCACTTCATGGCC 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4896 GACTGGAAGAACGCCTGTGCCATGCTGAAGGACGGGACTGCAGGCTCGCACTTCATGGCC 4955

Qy        721 TCTCCGCAGTGTGTGGGCTACAGCCGCAGCACAGCGGCCCCCCTCACGATGACTATGTGC 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4956 TCTCCGCAGTGTGTGGGCTACAGCCGCAGCACAGCGGCCCCCCTCACGATGACTATGTGC 5015

Qy        781 CTGCCTGACCTGAAGGAGATCCAGAGGGCTGTGAAGCTCTGGGTGAGGTCGCTGGATGCC 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5016 CTGCCTGACCTGAAGGAGATCCAGAGGGCTGTGAAGCTCTGGGTGAGGTCGCTGGATGCC 5075

Qy        841 CAGTCGGTCTACGTTGCTACTGATTCCGAGAGTTATGTGCCTGAGCTCCAACAGCTCTTC 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5076 CAGTCGGTCTACGTTGCTACTGATTCCGAGAGTTATGTGCCTGAGCTCCAACAGCTCTTC 5135

Qy        901 AAAGGGAAGGTGAAGGTGGTGAGCCTGAAGCCTGAGGTGGCCCAGGTCGACCTGTACATC 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5136 AAAGGGAAGGTGAAGGTGGTGAGCCTGAAGCCTGAGGTGGCCCAGGTCGACCTGTACATC 5195

Qy        961 CTCGGCCAAGCCGACCACTTTATTGGCAACTGTGTCTCCTCCTTCACTGCCTTTGTGAAG 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5196 CTCGGCCAAGCCGACCACTTTATTGGCAACTGTGTCTCCTCCTTCACTGCCTTTGTGAAG 5255

Qy       1021 CGGGAGCGGGACCTCCAGGGGAGGCCGTCTTCTTTCTTCGGCATGGACAGGCCCCCTAAG 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5256 CGGGAGCGGGACCTCCAGGGGAGGCCGTCTTCTTTCTTCGGCATGGACAGGCCCCCTAAG 5315

Qy       1081 CTGCGGGACGAGTTCTGATT 1100
              ||||||||||||||||||||
Db       5316 CTGCGGGACGAGTTCTGATT 5335


RESULT 9
D80002
LOCUS       D80002                 5189 bp    mRNA    linear   PRI 10-JAN-2003
DEFINITION  Homo sapiens KIAA0180 mRNA, complete cds.
ACCESSION   D80002
VERSION     D80002.2  GI:20521837
KEYWORDS    .
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1
  AUTHORS   Nagase,T., Seki,N., Ishikawa,K., Tanaka,A. and Nomura,N.
  TITLE     Prediction of the coding sequences of unidentified human genes. V.
```

```
            The coding sequences of 40 new genes (KIAA0161-KIAA0200) deduced by
            analysis of cDNA clones from human cell line KG-1
  JOURNAL   DNA Res. 3 (1), 17-24 (1996)
  MEDLINE   96281124
   PUBMED   8724849
REFERENCE   2
  AUTHORS   Chiang,P.W., Wang,S., Smithivas,P., Song,W.J., Ramamoorthy,S.,
            Hillman,J., Puett,S., Van Keuren,M.L., Crombez,E., Kumar,A.,
            Glover,T.W., Miller,D.E., Tsai,C.H., Blackburn,C.C., Chen,X.N.,
            Sun,Z., Cheng,J.F., Korenberg,J.R. and Kurnit,D.M.
  TITLE     Identification and analysis of the human and murine putative
            chromatin structure regulator SUPT6H and Supt6h
  JOURNAL   Genomics 34 (3), 328-333 (1996)
  MEDLINE   96374824
   PUBMED   8786132
REFERENCE   3  (bases 1 to 5189)
  AUTHORS   Ohara,O., Nagase,T., Kikuno,R. and Nomura,N.
  TITLE     Direct Submission
  JOURNAL   Submitted (12-DEC-1995) Osamu Ohara, Kazusa DNA Research Institute;
            1532-3, Yana, Kisarazu, Chiba 292-0812, Japan
            (E-mail:cdnainfo@kazusa.or.jp, Tel:+81-438-52-3913)
COMMENT     On May 9, 2002 this sequence version replaced gi:1136419.
FEATURES             Location/Qualifiers
     source          1. .5189
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="ha02567s1"
                     /sex="male"
                     /cell_line="KG-1"
                     /cell_type="myeloblast"
                     /tissue_type="brain"
                     /clone_lib="pBluescriptII SK plus"
                     /note="This sequence was obtained by subcloning of the DNA
                     fragments derived from two cDNA clones (1 - 1066 was
                     derived from pf00933 and 1067 - 5189 was derived from
                     ha02567)."
     gene            1. .5189
                     /gene="KIAA0180"
     CDS             <1. .1213
                     /gene="KIAA0180"
                     /citation=[2]
                     /codon_start=2
                     /protein_id="BAA11497.2"
                     /db_xref="GI:20521838"
                     /translation="PPRLCAAAGSGSRADMGAAAWARPLSVSFLLLLLPLPGMPAGSW
                     DPAGYLLYCPCMGRFGNQADHFLGSLAFAKLLNRTLAVPPWIEYQHHKPPFTNLHVSY
                     QKYFKLEPLQAYHRVISLEDFMEKLAPTHWPPEKRVAYCFEVAAQRSPDKKTCPMKEG
                     NPFGPFWDQFHVSFNKSELFTGISFSASYREQWSQRFSPKEHPVLALPGAPAQFPVLE
                     EHRPLQKYMVWSDEMVKTGEAQIHAHLVRPYVGIHLRIGSDWKNACAMLKDGTAGSHF
                     MASPQCVGYSRSTAAPLTMTMCLPDLKEIQRAVKLWVRSLDAQSVYVATDSESYVPEL
                     QQLFKGKVKVVSLKPEVAQVDLYILGQADHFIGNCVSSFTAFVKRERDLQGRPSSFFG
                     MDRPPKLRDEF"
ORIGIN

  Query Match             99.7%;  Score 1096.8;  DB 9;  Length 5189;
  Best Local Similarity   99.8%;  Pred. No. 5.9e-253;
  Matches 1098;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy          1 ATGCCCGCGGGCTCCTGGGACCCGGCCGGTTACCTGCTCTACTGCCCCTGCATGGGGCGC 60
              ||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        116 ATGCCTGCGGGCTCCTGGGACCCGGCCGGTTACCTGCTCTACTGCCCCTGCATGGGGCGC 175

Qy         61 TTTGGGAACCAGGCCGATCACTTCTTGGGCTCTCTGGCATTTGCAAAGCTGCTAAACCGT 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        176 TTTGGGAACCAGGCCGATCACTTCTTGGGCTCTCTGGCATTTGCAAAGCTGCTAAACCGT 235

Qy        121 ACCTTGGCTGTCCCTCCTTGGATTGAGTACCAGCATCACAAGCCTCCTTTCACCAACCTC 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        236 ACCTTGGCTGTCCCTCCTTGGATTGAGTACCAGCATCACAAGCCTCCTTTCACCAACCTC 295

Qy        181 CATGTGTCCTACCAGAAGTACTTCAAGCTGGAGCCCCTCCAGGCTTACCATCGGGTCATC 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

Db        296 CATGTGTCCTACCAGAAGTACTTCAAGCTGGAGCCCCTCCAGGCTTACCATCGGGTCATC 355

Qy        241 AGCTTGGAGGATTTCATGGAGAAGCTGGCACCCACCCACTGGCCCCCTGAGAAGCGGGTG 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        356 AGCTTGGAGGATTTCATGGAGAAGCTGGCACCCACCCACTGGCCCCCTGAGAAGCGGGTG 415

Qy        301 GCATACTGCTTTGAGGTGGCAGCCCAGCGAAGCCCAGATAAGAAGACGTGCCCCATGAAG 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        416 GCATACTGCTTTGAGGTGGCAGCCCAGCGAAGCCCAGATAAGAAGACGTGCCCCATGAAG 475

Qy        361 GAAGGAAACCCCTTTGGCCCATTCTGGGATCAGTTTCATGTGAGTTTCAACAAGTCGGAG 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        476 GAAGGAAACCCCTTTGGCCCATTCTGGGATCAGTTTCATGTGAGTTTCAACAAGTCGGAG 535

Qy        421 CTTTTTACAGGCATTTCCTTCAGTGCTTCCTACAGAGAACAATGGAGCCAGAGATTTTCT 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        536 CTTTTTACAGGCATTTCCTTCAGTGCTTCCTACAGAGAACAATGGAGCCAGAGATTTTCT 595

Qy        481 CCAAAGGAACATCCGGTGCTTGCCCTGCCAGGAGCCCCAGCCCAGTTCCCCGTCCTAGAA 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        596 CCAAAGGAACATCCGGTGCTTGCCCTGCCAGGAGCCCCAGCCCAGTTCCCCGTCCTAGAG 655

Qy        541 GAACACAGGCCACTACAGAAGTACATGGTATGGTCAGACGAAATGGTGAAGACGGGAGAG 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        656 GAACACAGGCCACTACAGAAGTACATGGTATGGTCAGACGAAATGGTGAAGACGGGAGAG 715

Qy        601 GCCCAGATTCATGCCCACCTTGTCCGGCCCTATGTGGGCATTCATCTGCGCATTGGCTCT 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        716 GCCCAGATTCATGCCCACCTTGTCCGGCCCTATGTGGGCATTCATCTGCGCATTGGCTCT 775

Qy        661 GACTGGAAGAACGCCTGTGCCATGCTGAAGGACGGGACTGCAGGCTCGCACTTCATGGCC 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        776 GACTGGAAGAACGCCTGTGCCATGCTGAAGGACGGGACTGCAGGCTCGCACTTCATGGCC 835

Qy        721 TCTCCGCAGTGTGTGGGCTACAGCCGCAGCACAGCGGCCCCCCTCACGATGACTATGTGC 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        836 TCTCCGCAGTGTGTGGGCTACAGCCGCAGCACAGCGGCCCCCCTCACGATGACTATGTGC 895

Qy        781 CTGCCTGACCTGAAGGAGATCCAGAGGGCTGTGAAGCTCTGGGTGAGGTCGCTGGATGCC 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        896 CTGCCTGACCTGAAGGAGATCCAGAGGGCTGTGAAGCTCTGGGTGAGGTCGCTGGATGCC 955

Qy        841 CAGTCGGTCTACGTTGCTACTGATTCCGAGAGTTATGTGCCTGAGCTCCAACAGCTCTTC 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        956 CAGTCGGTCTACGTTGCTACTGATTCCGAGAGTTATGTGCCTGAGCTCCAACAGCTCTTC 1015

Qy        901 AAAGGGAAGGTGAAGGTGGTGAGCCTGAAGCCTGAGGTGGCCCAGGTCGACCTGTACATC 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1016 AAAGGGAAGGTGAAGGTGGTGAGCCTGAAGCCTGAGGTGGCCCAGGTCGACCTGTACATC 1075

Qy        961 CTCGGCCAAGCCGACCACTTTATTGGCAACTGTGTCTCCTCCTTCACTGCCTTTGTGAAG 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1076 CTCGGCCAAGCCGACCACTTTATTGGCAACTGTGTCTCCTCCTTCACTGCCTTTGTGAAG 1135

Qy       1021 CGGGAGCGGGACCTCCAGGGGAGGCCGTCTTCTTTCTTCGGCATGGACAGGCCCCCTAAG 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1136 CGGGAGCGGGACCTCCAGGGGAGGCCGTCTTCTTTCTTCGGCATGGACAGGCCCCCTAAG 1195

Qy       1081 CTGCGGGACGAGTTCTGATT 1100
              ||||||||||||||||||||
Db       1196 CTGCGGGACGAGTTCTGATT 1215


RESULT 10
CQ727777
LOCUS       CQ727777               1167 bp    DNA     linear   PAT 03-FEB-2004
DEFINITION  Sequence 13711 from Patent WO02068579.
ACCESSION   CQ727777
VERSION     CQ727777.1  GI:42294771
KEYWORDS    .
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
```

```
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1
  AUTHORS   Venter,C.J., Adams,M.C., Li,P.W. and Myers,E.W.
  TITLE     Kits, such as nucleic acid arrays, comprising a majority of
            humanexons or transcripts, for detecting expression and other uses
            thereof
  JOURNAL   Patent: WO 02068579-A 13711 06-SEP-2002;
            PE Corporation (NY) (US)
FEATURES             Location/Qualifiers
     source          1. .1167
                     /organism="Homo sapiens"
                     /mol_type="unassigned DNA"
                     /db_xref="taxon:9606"
ORIGIN

  Query Match             99.7%;  Score 1096.4;  DB 6;  Length 1167;
  Best Local Similarity   99.9%;  Pred. No. 7.3e-253;
  Matches 1097;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 ATGCCCGCGGGCTCCTGGGACCCGGCCGGTTACCTGCTCTACTGCCCCTGCATGGGGCGC 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         70 ATGCCCGCGGGCTCCTGGGACCCGGCCGGTTACCTGCTCTACTGCCCCTGCATGGGGCGC 129

Qy         61 TTTGGGAACCAGGCCGATCACTTCTTGGGCTCTCTGGCATTTGCAAAGCTGCTAAACCGT 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        130 TTTGGGAACCAGGCCGATCACTTCTTGGGCTCTCTGGCATTTGCAAAGCTGCTAAACCGT 189

Qy        121 ACCTTGGCTGTCCCTCCTTGGATTGAGTACCAGCATCACAAGCCTCCTTTCACCAACCTC 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        190 ACCTTGGCTGTCCCTCCTTGGATTGAGTACCAGCATCACAAGCCTCCTTTCACCAACCTC 249

Qy        181 CATGTGTCCTACCAGAAGTACTTCAAGCTGGAGCCCCTCCAGGCTTACCATCGGGTCATC 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        250 CATGTGTCCTACCAGAAGTACTTCAAGCTGGAGCCCCTCCAGGCTTACCATCGGGTCATC 309

Qy        241 AGCTTGGAGGATTTCATGGAGAAGCTGGCACCCACCCACTGGCCCCCTGAGAAGCGGGTG 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        310 AGCTTGGAGGATTTCATGGAGAAGCTGGCACCCACCCACTGGCCCCCTGAGAAGCGGGTG 369

Qy        301 GCATACTGCTTTGAGGTGGCAGCCCAGCGAAGCCCAGATAAGAAGACGTGCCCCATGAAG 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        370 GCATACTGCTTTGAGGTGGCAGCCCAGCGAAGCCCAGATAAGAAGACGTGCCCCATGAAG 429

Qy        361 GAAGGAAACCCCTTTGGCCCATTCTGGGATCAGTTTCATGTGAGTTTCAACAAGTCGGAG 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        430 GAAGGAAACCCCTTTGGCCCATTCTGGGATCAGTTTCATGTGAGTTTCAACAAGTCGGAG 489

Qy        421 CTTTTTACAGGCATTTCCTTCAGTGCTTCCTACAGAGAACAATGGAGCCAGAGATTTTCT 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        490 CTTTTTACAGGCATTTCCTTCAGTGCTTCCTACAGAGAACAATGGAGCCAGAGATTTTCT 549

Qy        481 CCAAAGGAACATCCGGTGCTTGCCCTGCCAGGAGCCCCAGCCCAGTTCCCCGTCCTAGAA 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        550 CCAAAGGAACATCCGGTGCTTGCCCTGCCAGGAGCCCCAGCCCAGTTCCCCGTCCTAGAG 609

Qy        541 GAACACAGGCCACTACAGAAGTACATGGTATGGTCAGACGAAATGGTGAAGACGGGAGAG 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        610 GAACACAGGCCACTACAGAAGTACATGGTATGGTCAGACGAAATGGTGAAGACGGGAGAG 669

Qy        601 GCCCAGATTCATGCCCACCTTGTCCGGCCCTATGTGGGCATTCATCTGCGCATTGGCTCT 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        670 GCCCAGATTCATGCCCACCTTGTCCGGCCCTATGTGGGCATTCATCTGCGCATTGGCTCT 729

Qy        661 GACTGGAAGAACGCCTGTGCCATGCTGAAGGACGGGACTGCAGGCTCGCACTTCATGGCC 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        730 GACTGGAAGAACGCCTGTGCCATGCTGAAGGACGGGACTGCAGGCTCGCACTTCATGGCC 789

Qy        721 TCTCCGCAGTGTGTGGGCTACAGCCGCAGCACAGCGGCCCCCCTCACGATGACTATGTGC 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        790 TCTCCGCAGTGTGTGGGCTACAGCCGCAGCACAGCGGCCCCCCTCACGATGACTATGTGC 849

Qy        781 CTGCCTGACCTGAAGGAGATCCAGAGGGCTGTGAAGCTCTGGGTGAGGTCGCTGGATGCC 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        850 CTGCCTGACCTGAAGGAGATCCAGAGGGCTGTGAAGCTCTGGGTGAGGTCGCTGGATGCC 909

Qy        841 CAGTCGGTCTACGTTGCTACTGATTCCGAGAGTTATGTGCCTGAGCTCCAACAGCTCTTC 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        910 CAGTCGGTCTACGTTGCTACTGATTCCGAGAGTTATGTGCCTGAGCTCCAACAGCTCTTC 969

Qy        901 AAAGGGAAGGTGAAGGTGGTGAGCCTGAAGCCTGAGGTGGCCCAGGTCGACCTGTACATC 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        970 AAAGGGAAGGTGAAGGTGGTGAGCCTGAAGCCTGAGGTGGCCCAGGTCGACCTGTACATC 1029

Qy        961 CTCGGCCAAGCCGACCACTTTATTGGCAACTGTGTCTCCTCCTTCACTGCCTTTGTGAAG 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1030 CTCGGCCAAGCCGACCACTTTATTGGCAACTGTGTCTCCTCCTTCACTGCCTTTGTGAAG 1089

Qy       1021 CGGGAGCGGGACCTCCAGGGGAGGCCGTCTTCTTTCTTCGGCATGGACAGGCCCCCTAAG 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1090 CGGGAGCGGGACCTCCAGGGGAGGCCGTCTTCTTTCTTCGGCATGGACAGGCCCCCTAAG 1149

Qy       1081 CTGCGGGACGAGTTCTGA 1098
              ||||||||||||||||||
Db       1150 CTGCGGGACGAGTTCTGA 1167


RESULT 11
AX780153
LOCUS       AX780153               5266 bp    DNA     linear   PAT 14-JUL-2003
DEFINITION  Sequence 2310 from Patent WO03039443.
ACCESSION   AX780153
VERSION     AX780153.1  GI:32697147
KEYWORDS    .
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1
  AUTHORS   Haferlach,T., Schoch,C., Kern,W., Kohlmann,A., Schnittger,S.,
            Dugas,M., Eils,R., Brors,B. and Mergenthaler,S.
  TITLE     Novel genetic markers for leukemias
  JOURNAL   Patent: WO 03039443-A 2310 15-MAY-2003;
            Deutsches Krebsforschungszentrum (DE) ;
            Ludwig-Maximilian-Universitaet Muenchen (DE) ; Haferlach, Torsten,
            PD Dr. Dr. (DE) ; Schoch, Claudia (DE) ; Kern, Wolfgang (DE)
FEATURES             Location/Qualifiers
     source          1. .5266
                     /organism="Homo sapiens"
                     /mol_type="unassigned DNA"
                     /db_xref="taxon:9606"
ORIGIN

  Query Match             99.6%;  Score 1095.8;  DB 6;  Length 5266;
  Best Local Similarity   99.7%;  Pred. No. 1e-252;
  Matches 1097;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy          1 ATGCCCGCGGGCTCCTGGGACCCGGCCGGTTACCTGCTCTACTGCCCCTGCATGGGGCGC 60
              ||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        132 ATGCCTGCGGGCTCCTGGGACCCGGCCGGTTACCTGCTCTACTGCCCCTGCATGGGGCGC 191

Qy         61 TTTGGGAACCAGGCCGATCACTTCTTGGGCTCTCTGGCATTTGCAAAGCTGCTAAACCGT 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        192 TTTGGGAACCAGGCCGATCACTTCTTGGGCTCTCTGGCATTTGCAAAGCTGCTAAACCGT 251

Qy        121 ACCTTGGCTGTCCCTCCTTGGATTGAGTACCAGCATCACAAGCCTCCTTTCACCAACCTC 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        252 ACCTTGGCTGTCCCTCCTTGGATTGAGTACCAGCATCACAAGCCTCCTTTCACCAACCTC 311

Qy        181 CATGTGTCCTACCAGAAGTACTTCAAGCTGGAGCCCCTCCAGGCTTACCATCGGGTCATC 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        312 CATGTGTCCTACCAGAAGTACTTCAAGCTGGAGCCCCTCCAGGCTTACCATCGGGTCATC 371

Qy        241 AGCTTGGAGGATTTCATGGAGAAGCTGGCACCCACCCACTGGCCCCCTGAGAAGCGGGTG 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        372 AGCTTGGAGGATTTCATGGAGAAGCTGGCACCCACCCACTGGCCCCCTGAGAAGCGGGTG 431
```

```
Qy        301 GCATACTGCTTTGAGGTGGCAGCCCAGCGAAGCCCAGATAAGAAGACGTGCCCCATGAAG 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        432 GCATACTGCTTTGAGGTGGCAGCCCAGCGAAGCCCAGATAAGAAGACGTGCCCCATGAAG 491

Qy        361 GAAGGAAACCCCTTTGGCCCATTCTGGGATCAGTTTCATGTGAGTTTCAACAAGTCGGAG 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        492 GAAGGAAACCCCTTTGGCCCATTCTGGGATCAGTTTCATGTGAGTTTCAACAAGTCGGAG 551

Qy        421 CTTTTTACAGGCATTTCCTTCAGTGCTTCCTACAGAGAACAATGGAGCCAGAGATTTTCT 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        552 CTTTTTACAGGCATTTCCTTCAGTGCTTCCTACAGAGAACAATGGAGCCAGAGATTTTCT 611

Qy        481 CCAAAGGAACATCCGGTGCTTGCCCTGCCAGGAGCCCCAGCCCAGTTCCCCGTCCTAGAA 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        612 CCAAAGGAACATCCGGTGCTTGCCCTGCCAGGAGCCCCAGCCCAGTTCCCCGTCCTAGAG 671

Qy        541 GAACACAGGCCACTACAGAAGTACATGGTATGGTCAGACGAAATGGTGAAGACGGGAGAG 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        672 GAACACAGGCCACTACAGAAGTACATGGTATGGTCAGACGAAATGGTGAAGACGGGAGAG 731

Qy        601 GCCCAGATTCATGCCCACCTTGTCCGGCCCTATGTGGGCATTCATCTGCGCATTGGCTCT 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        732 GCCCAGATTCATGCCCACCTTGTCCGGCCCTATGTGGGCATTCATCTGCGCATTGGCTCT 791

Qy        661 GACTGGAAGAACGCCTGTGCCATGCTGAAGGACGGGACTGCAGGCTCGCACTTCATGGCC 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        792 GACTGGAAGAACGCCTGTGCCATGCTGAAGGACGGGACTGCAGGCTCGCACTTCATGGCC 851

Qy        721 TCTCCGCAGTGTGTGGGCTACAGCCGCAGCACAGCGGCCCCCCTCACGATGACTATGTGC 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        852 TCTCCGCAGTGTGTGGGCTACAGCCGCAGCACAGCGGCCCCCCTCACGATGACTATGTGC 911

Qy        781 CTGCCTGACCTGAAGGAGATCCAGAGGGCTGTGAAGCTCTGGGTGAGGTCGCTGGATGCC 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        912 CTGCCTGACCTGAAGGAGATCCAGAGGGCTGTGAAGCTCTGGGTGAGGTCGCTGGATGCC 971

Qy        841 CAGTCGGTCTACGTTGCTACTGATTCCGAGAGTTATGTGCCTGAGCTCCAACAGCTCTTC 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        972 CAGTCGGTCTACGTTGCTACTGATTCCGAGAGTTATGTGCCTGAGCTCCAACAGCTCTTC 1031

Qy        901 AAAGGGAAGGTGAAGGTGGTGAGCCTGAAGCCTGAGGTGGCCCAGGTCGACCTGTACATC 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1032 AAAGGGAAGGTGAAGGTGGTGAGCCTGAAGCCTGAGGTGGCCCAGGTCGACCTGTACATC 1091

Qy        961 CTCGGCCAAGCCGACCACTTTATTGGCAACTGTGTCTCCTCCTTCACTGCCTTTGTGAAG 1020
              ||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1092 CTCGGNCAAGCCGACCACTTTATTGGCAACTGTGTCTCCTCCTTCACTGCCTTTGTGAAG 1151

Qy       1021 CGGGAGCGGGACCTCCAGGGGAGGCCGTCTTCTTTCTTCGGCATGGACAGGCCCCCTAAG 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1152 CGGGAGCGGGACCTCCAGGGGAGGCCGTCTTCTTTCTTCGGCATGGACAGGCCCCCTAAG 1211

Qy       1081 CTGCGGGACGAGTTCTGATT 1100
              ||||||||||||||||||||
Db       1212 CTGCGGGACGAGTTCTGATT 1231
```

RESULT 12
HSM803330

```
LOCUS       HSM803330              4560 bp    mRNA    linear   PRI 13-MAY-2003
DEFINITION  Homo sapiens mRNA; cDNA DKFZp451J114 (from clone DKFZp451J114).
ACCESSION   AL832023
VERSION     AL832023.1  GI:21732563
KEYWORDS    .
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 4560)
  AUTHORS   Wambutt,R., Heubner,D., Mewes,H.W., Weil,B., Amid,C., Osanger,A.,
            Fobo,G., Han,M. and Wiemann,S.
  TITLE     Direct Submission
  JOURNAL   Submitted (13-MAY-2003) MIPS, Ingolstaedter Landstr.1, D-85764
            Neuherberg, GERMANY
COMMENT     Clone from S. Wiemann, Molecular Genome Analysis, German Cancer
            Research Center (DKFZ); Email s.wiemann@dkfz-heidelberg.de;
            sequenced by AGOWA (Berlin/Germany) within the cDNA sequencing
            consortium of the German Genome Project.
            This clone (DKFZp451J114) is available at the RZPD in Berlin.
            Please contact the RZPD: Ressourcenzentrum, Heubnerweg 6, 14059
            Berlin-Charlottenburg, GERMANY; Email: clone@rzpd.de Further
            information about the clone and the sequencing project is available
            at http://mips.gsf.de/proj/cDNA/.
FEATURES             Location/Qualifiers
     source          1. .4560
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="RZPD:DKFZp451J114"
                     /db_xref="taxon:9606"
                     /clone="DKFZp451J114"
                     /tissue_type="human skeletal muscle"
                     /clone_lib="451 (synonym: hlcc1). Vector pSport1; host
                     DH10B; sites NotI + SalI"
                     /dev_stage="adult"
     polyA_site      4538
ORIGIN
```

Query Match 99.6%; Score 1095.2; DB 9; Length 4560;
Best Local Similarity 99.7%; Pred. No. 1.4e-252;
Matches 1097; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

```
Qy          1 ATGCCCGCGGGCTCCTGGGACCCGGCCGGTTACCTGCTCTACTGCCCCTGCATGGGGCGC 60
              ||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         51 ATGCCTGCGGGCTCCTGGGACCCGGCCGGTTACCTGCTCTACTGCCCCTGCATGGGGCGC 110

Qy         61 TTTGGGAACCAGGCCGATCACTTCTTGGGCTCTCTGGCATTTGCAAAGCTGCTAAACCGT 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        111 TTTGGGAACCAGGCCGATCACTTCTTGGGCTCTCTGGCATTTGCAAAGCTGCTAAACCGT 170

Qy        121 ACCTTGGCTGTCCCTCCTTGGATTGAGTACCAGCATCACAAGCCTCCTTTCACCAACCTC 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        171 ACCTTGGCTGTCCCTCCTTGGATTGAGTACCAGCATCACAAGCCTCCTTTCACCAACCTC 230

Qy        181 CATGTGTCCTACCAGAAGTACTTCAAGCTGGAGCCCCTCCAGGCTTACCATCGGGTCATC 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        231 CATGTGTCCTACCAGAAGTACTTCAAGCTGGAGCCCCTCCAGGCTTACCATCGGGTCATC 290

Qy        241 AGCTTGGAGGATTTCATGGAGAAGCTGGCACCCACCCACTGGCCCCCTGAGAAGCGGGTG 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        291 AGCTTGGAGGATTTCATGGAGAAGCTGGCACCCACCCACTGGCCCCCTGAGAAGCGGGTG 350

Qy        301 GCATACTGCTTTGAGGTGGCAGCCCAGCGAAGCCCAGATAAGAAGACGTGCCCCATGAAG 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        351 GCATACTGCTTTGAGGTGGCAGCCCAGCGAAGCCCAGATAAGAAGACGTGCCCCATGAAG 410

Qy        361 GAAGGAAACCCCTTTGGCCCATTCTGGGATCAGTTTCATGTGAGTTTCAACAAGTCGGAG 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        411 GAAGGAAACCCCTTTGGCCCATTCTGGGATCAGTTTCATGTGAGTTTCAACAAGTCGGAG 470

Qy        421 CTTTTTACAGGCATTTCCTTCAGTGCTTCCTACAGAGAACAATGGAGCCAGAGATTTTCT 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        471 CTTTTTACAGGCATTTCCTTCAGTGCTTCCTACAGAGAACAATGGAGCCAGAGATTTTCT 530

Qy        481 CCAAAGGAACATCCGGTGCTTGCCCTGCCAGGAGCCCCAGCCCAGTTCCCCGTCCTAGAA 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        531 CCAAAGGAACATCCGGTGCTTGCCCTGCCAGGAGCCCCAGCCCAGTTCCCCGTCCTAGAG 590

Qy        541 GAACACAGGCCACTACAGAAGTACATGGTATGGTCAGACGAAATGGTGAAGACGGGAGAG 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        591 GAACACAGGCCACTACAGAAGTACATGGTATGGTCAGACGAAATGGTGAAGACGGGAGAG 650

Qy        601 GCCCAGATTCATGCCCACCTTGTCCGGCCCTATGTGGGCATTCATCTGCGCATTGGCTCT 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        651 GCCCAGATTCATGCCCACCTTGTCCGGCCCTATGTGGGCATTCATCTGCGCATTGGCTCT 710

Qy        661 GACTGGAAGAACGCCTGTGCCATGCTGAAGGACGGGACTGCAGGCTCGCACTTCATGGCC 720
```

```
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        711 GACTGGAAGAACGCCTGTGCCATGCTGAAGGACGGGACTGCAGGCTCGCACTTCATGGCC 770

Qy        721 TCTCCGCAGTGTGTGGGCTACAGCCGCAGCACAGCGGCCCCCCTCACGATGACTATGTGC 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        771 TCTCCGCAGTGTGTGGGCTACAGCCGCAGCACAGCGGCCCCCCTCACGATGACTATGTGC 830

Qy        781 CTGCCTGACCTGAAGGAGATCCAGAGGGCTGTGAAGCTCTGGGTGAGGTCGCTGGATGCC 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        831 CTGCCTGACCTGAAGGAGATCCAGAGGGCTGTGAAGCTCTGGGTGAGGTCGCTGGATGCC 890

Qy        841 CAGTCGGTCTACGTTGCTACTGATTCCGAGAGTTATGTGCCTGAGCTCCAACAGCTCTTC 900
              ||||||||||||||||||||| ||||||||||||||||||||||||||||||||||||||
Db        891 CAGTCGGTCTACGTTGCTACCGATTCCGAGAGTTATGTGCCTGAGCTCCAACAGCTCTTC 950

Qy        901 AAAGGGAAGGTGAAGGTGGTGAGCCTGAAGCCTGAGGTGGCCCAGGTCGACCTGTACATC 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        951 AAAGGGAAGGTGAAGGTGGTGAGCCTGAAGCCTGAGGTGGCCCAGGTCGACCTGTACATC 1010

Qy        961 CTCGGCCAAGCCGACCACTTTATTGGCAACTGTGTCTCCTCCTTCACTGCCTTTGTGAAG 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1011 CTCGGCCAAGCCGACCACTTTATTGGCAACTGTGTCTCCTCCTTCACTGCCTTTGTGAAG 1070

Qy       1021 CGGGAGCGGGACCTCCAGGGGAGGCCGTCTTCTTTCTTCGGCATGGACAGGCCCCCTAAG 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1071 CGGGAGCGGGACCTCCAGGGGAGGCCGTCTTCTTTCTTCGGCATGGACAGGCCCCCTAAG 1130

Qy       1081 CTGCGGGACGAGTTCTGATT 1100
              ||||||||||||||||||||
Db       1131 CTGCGGGACGAGTTCTGATT 1150
```

```
RESULT 13
AJ781500
LOCUS       AJ781500                1167 bp    mRNA    linear   PRI 07-JUL-2004
DEFINITION  Pan troglodytes mRNA for protein-O-fucosyltransferase 1 (fut12
            gene).
ACCESSION   AJ781500
VERSION     AJ781500.1  GI:50057069
KEYWORDS    fut12 gene; pofut1 gene; protein-O-fucosyltransferase 1.
SOURCE      Pan troglodytes (chimpanzee)
  ORGANISM  Pan troglodytes
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Pan.
REFERENCE   1
  AUTHORS   Martinez-Duncker,I., Mollicone,R., Candelier,J.J., Breton,C. and
            Oriol,R.
  TITLE     A new superfamily of protein-O-fucosyltransferases,
            alpha2-fucosyltransferases, and alpha6-fucosyltransferases:
            phylogeny and identification of conserved peptide motifs
  JOURNAL   Glycobiology 13 (12), 1C-5C (2003)
   PUBMED   12966037
REFERENCE   2
  AUTHORS   Martinez-Duncker,I., Oriol,R. and Mollicone,R.
  TITLE     Phylogeny of fucosyltransferases
  JOURNAL   Unpublished
REFERENCE   3  (bases 1 to 1167)
  AUTHORS   Oriol,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (06-JUL-2004) Oriol R., U504, Inserm, 16 Av. Paul
            Vaillant-Couturier, 94807, FRANCE
FEATURES             Location/Qualifiers
     source          1..1167
                     /organism="Pan troglodytes"
                     /mol_type="mRNA"
                     /db_xref="taxon:9598"
     gene            1..1167
                     /gene="fut12"
     CDS             1..1167
                     /gene="fut12"
                     /function="adds fucose to Ser or Thr"
                     /note="pofut1"
                     /codon_start=1
                     /product="protein-O-fucosyltransferase 1"
                     /protein_id="CAH03712.1"
                     /db_xref="GI:50057070"
                     /translation="MGAAAWARPLSVSFLLLLLPLPGMPAGSWDPAGYLLYCPCMGRF
                     GNQADHFLGSLAFAKLLNRTLAVPPWIEYQHHKPPFTNLHVSYQKYFKLEPLQAYHRV
                     ISLEDFMEKLAPTHWPPEKRVAYCFEVAAQRSPDKKTCPMKEGNPFGPFWDQFHVSFN
                     KSELFTGISFSASYREQWSQRFSPKEHPVLALPGAPAQFPVLEEHRPLQKYMVWSDEM
                     VKTGEAQIHAHLVRPYVGIHLRIGSDWKNACAMLKDGTAGSHFMASPQCVGYSRSTAA
                     PLTMTMCLPDLKEIQRAVKLWVRSLDAQSVYVATDSESYVPELQQLFKGKVKVVSLKP
                     EVAQVDLYILGQADHFIGNCVSSFTAFVKRERDLQGRPSSFFGMDRPPKLRDEF"
ORIGIN

  Query Match             98.8%;  Score 1086.8;  DB 9;  Length 1167;
  Best Local Similarity   99.4%;  Pred. No. 1.5e-250;
  Matches 1091;  Conservative    0;  Mismatches    7;  Indels    0;  Gaps    0;

Qy          1 ATGCCCGCGGGCTCCTGGGACCCGGCCGGTTACCTGCTCTACTGCCCCTGCATGGGGCGC 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         70 ATGCCCGCGGGCTCCTGGGACCCGGCCGGTTACCTGCTCTACTGCCCCTGCATGGGGCGC 129

Qy         61 TTTGGGAACCAGGCCGATCACTTCTTGGGCTCTCTGGCATTTGCAAAGCTGCTAAACCGT 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        130 TTTGGGAACCAGGCCGATCACTTCTTGGGCTCTCTGGCATTTGCAAAGCTGCTAAACCGT 189

Qy        121 ACCTTGGCTGTCCCTCCTTGGATTGAGTACCAGCATCACAAGCCTCCTTTCACCAACCTC 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        190 ACCTTGGCTGTCCCTCCTTGGATTGAGTACCAGCATCACAAGCCTCCTTTCACCAACCTC 249

Qy        181 CATGTGTCCTACCAGAAGTACTTCAAGCTGGAGCCCCTCCAGGCTTACCATCGGGTCATC 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        250 CATGTGTCCTACCAGAAGTACTTCAAGCTGGAGCCCCTCCAGGCTTACCATCGGGTCATC 309

Qy        241 AGCTTGGAGGATTTCATGGAGAAGCTGGCACCCACCCACTGGCCCCCTGAGAAGCGGGTG 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        310 AGCTTGGAGGATTTCATGGAGAAGCTGGCACCCACCCACTGGCCCCCTGAGAAGCGGGTG 369

Qy        301 GCATACTGCTTTGAGGTGGCAGCCCAGCGAAGCCCAGATAAGAAGACGTGCCCCATGAAG 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        370 GCATACTGCTTTGAGGTGGCAGCCCAGCGAAGCCCAGATAAGAAGACGTGCCCCATGAAG 429

Qy        361 GAAGGAAACCCCTTTGGCCCATTCTGGGATCAGTTTCATGTGAGTTTCAACAAGTCGGAG 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        430 GAAGGAAACCCCTTTGGCCCATTCTGGGATCAGTTTCATGTGAGTTTCAACAAGTCGGAG 489

Qy        421 CTTTTTACAGGCATTTCCTTCAGTGCTTCCTACAGAGAACAATGGAGCCAGAGATTTTCT 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        490 CTTTTTACAGGCATTTCCTTCAGTGCTTCCTACAGAGAACAATGGAGCCAGAGATTTTCT 549

Qy        481 CCAAAGGAACATCCGGTGCTTGCCCTGCCAGGAGCCCCAGCCCAGTTCCCCGTCCTAGAA 540
              |||||||||||||||||||||||||||||||||||||| ||||||||||||||||| || 
Db        550 CCAAAGGAACATCCGGTGCTTGCCCTGCCAGGAGCCCCGGCCCAGTTCCCCGTCCTGGAG 609

Qy        541 GAACACAGGCCACTACAGAAGTACATGGTATGGTCAGACGAAATGGTGAAGACGGGAGAG 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        610 GAACACAGGCCACTACAGAAGTACATGGTATGGTCAGACGAAATGGTGAAGACGGGAGAG 669

Qy        601 GCCCAGATTCATGCCCACCTTGTCCGGCCCTATGTGGGCATTCATCTGCGCATTGGCTCT 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        670 GCCCAGATTCATGCCCACCTTGTCCGGCCCTATGTGGGCATTCATCTGCGCATTGGCTCT 729

Qy        661 GACTGGAAGAACGCCTGTGCCATGCTGAAGGACGGGACTGCAGGCTCGCACTTCATGGCC 720
              ||||||||||||||||||||||||||||||||||||| ||||||||||||||||||||||
Db        730 GACTGGAAGAACGCCTGTGCCATGCTGAAGGACGGGACCGCAGGCTCGCACTTCATGGCC 789

Qy        721 TCTCCGCAGTGTGTGGGCTACAGCCGCAGCACAGCGGCCCCCCTCACGATGACTATGTGC 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        790 TCTCCGCAGTGTGTGGGCTACAGCCGCAGCACAGCGGCCCCCCTCACGATGACTATGTGC 849

Qy        781 CTGCCTGACCTGAAGGAGATCCAGAGGGCTGTGAAGCTCTGGGTGAGGTCGCTGGATGCC 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        850 CTGCCTGACCTGAAGGAGATCCAGAGGGCTGTGAAGCTCTGGGTGAGGTCGCTGGATGCC 909

Qy        841 CAGTCGGTCTACGTTGCTACTGATTCCGAGAGTTATGTGCCTGAGCTCCAACAGCTCTTC 900
```

```
Db        910 CAGTCGGTCTATGTTGCTACTGATTCCGAGAGTTATGTGCCTGAGCTCCAACAGCTCTTC 969

Qy        901 AAAGGGAAGGTGAAGGTGGTGAGCCTGAAGCCTGAGGTGGCCCAGGTCGACCTGTACATC 960

Db        970 AAAGGGAAGGTGAAGGTGGTGAGCCTGAAGCCTGAGGTGGCCCAGGTTGACCTGTACATC 1029

Qy        961 CTCGGCCAAGCCGACCACTTTATTGGCAACTGTGTCTCCTCCTTCACTGCCTTTGTGAAG 1020

Db       1030 CTCGGCCAAGCTGACCACTTTATTGGCAACTGTGTCTCCTCCTTCACTGCCTTTGTGAAG 1089

Qy       1021 CGGGAGCGGGACCTCCAGGGGAGGCCGTCTTCTTTCTTCGGCATGGACAGGCCCCCTAAG 1080

Db       1090 CGGGAGCGGGACCTCCAGGGGAGGCCGTCTTCTTTCTTCGGCATGGACAGGCCCCCTAAG 1149

Qy       1081 CTGCGGGACGAGTTCTGA 1098

Db       1150 CTGCGGGACGAGTTCTGA 1167


RESULT 14
AR163453
LOCUS       AR163453               5009 bp    DNA     linear   PAT 17-OCT-2001
DEFINITION  Sequence 8 from patent US 6270987.
ACCESSION   AR163453
VERSION     AR163453.1  GI:16234058
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 5009)
  AUTHORS   Wang,Y. and Spellman,M.W.
  TITLE     O-fucosyltransferase
  JOURNAL   Patent: US 6270987-A 8 07-AUG-2001;
FEATURES             Location/Qualifiers
     source          1. .5009
                     /organism="unknown"
                     /mol_type="unassigned DNA"
ORIGIN

  Query Match           93.9%;  Score 1033.4;  DB 6;  Length 5009;
  Best Local Similarity   99.9%;  Pred. No. 1e-237;
  Matches 1034;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy         66 GAACCAGGCCGATCACTTCTTGGGCTCTCTGGCATTTGCAAAGCTGCTAAACCGTACCTT 125

Db          1 GAACCAGGCCGATCACTTCTTGGGCTCTCTGGCATTTGCAAAGCTGCTAAACCGTACCTT 60

Qy        126 GGCTGTCCCTCCTTGGATTGAGTACCAGCATCACAAGCCTCCTTTCACCAACCTCCATGT 185

Db         61 GGCTGTCCCTCCTTGGATTGAGTACCAGCATCACAAGCCTCCTTTCACCAACCTCCATGT 120

Qy        186 GTCCTACCAGAAGTACTTCAAGCTGGAGCCCCTCCAGGCTTACCATCGGGTCATCAGCTT 245

Db        121 GTCCTACCAGAAGTACTTCAAGCTGGAGCCCCTCCAGGCTTACCATCGGGTCATCAGCTT 180

Qy        246 GGAGGATTTCATGGAGAAGCTGGCACCCACCCACTGGCCCCCTGAGAAGCGGGTGGCATA 305

Db        181 GGAGGATTTCATGGAGAAGCTGGCACCCACCCACTGGCCCCCTGAGAAGCGGGTGGCATA 240

Qy        306 CTGCTTTGAGGTGGCAGCCCAGCGAAGCCCAGATAAGAAGACGTGCCCCATGAAGGAAGG 365

Db        241 CTGCTTTGAGGTGGCAGCCCAGCGAAGCCCAGATAAGAAGACGTGCCCCATGAAGGAAGG 300

Qy        366 AAACCCCTTTGGCCCATTCTGGGATCAGTTTCATGTGAGTTTCAACAAGTCGGAGCTTTT 425

Db        301 AAACCCCTTTGGCCCATTCTGGGATCAGTTTCATGTGAGTTTCAACAAGTCGGAGCTTTT 360

Qy        426 TACAGGCATTTCCTTCAGTGCTTCCTACAGAGAACAATGGAGCCAGAGATTTTCTCCAAA 485

Db        361 TACAGGCATTTCCTTCAGTGCTTCCTACAGAGAACAATGGAGCCAGAGATTTTCTCCAAA 420

Qy        486 GGAACATCCGGTGCTTGCCCTGCCAGGAGCCCCAGCCCAGTTCCCCGTCCTAGAAGAACA 545

Db        421 GGAACATCCGGTGCTTGCCCTGCCAGGAGCCCCAGCCCAGTTCCCCGTCCTAGAGGAACA 480

Qy        546 CAGGCCACTACAGAAGTACATGGTATGGTCAGACGAAATGGTGAAGACGGGAGAGGCCCA 605

Db        481 CAGGCCACTACAGAAGTACATGGTATGGTCAGACGAAATGGTGAAGACGGGAGAGGCCCA 540

Qy        606 GATTCATGCCCACCTTGTCCGGCCCTATGTGGGCATTCATCTGCGCATTGGCTCTGACTG 665

Db        541 GATTCATGCCCACCTTGTCCGGCCCTATGTGGGCATTCATCTGCGCATTGGCTCTGACTG 600

Qy        666 GAAGAACGCCTGTGCCATGCTGAAGGACGGGACTGCAGGCTCGCACTTCATGGCCTCTCC 725

Db        601 GAAGAACGCCTGTGCCATGCTGAAGGACGGGACTGCAGGCTCGCACTTCATGGCCTCTCC 660

Qy        726 GCAGTGTGTGGGCTACAGCCGCAGCACAGCGGCCCCCCTCACGATGACTATGTGCCTGCC 785

Db        661 GCAGTGTGTGGGCTACAGCCGCAGCACAGCGGCCCCCCTCACGATGACTATGTGCCTGCC 720

Qy        786 TGACCTGAAGGAGATCCAGAGGGCTGTGAAGCTCTGGGTGAGGTCGCTGGATGCCCAGTC 845

Db        721 TGACCTGAAGGAGATCCAGAGGGCTGTGAAGCTCTGGGTGAGGTCGCTGGATGCCCAGTC 780

Qy        846 GGTCTACGTTGCTACTGATTCCGAGAGTTATGTGCCTGAGCTCCAACAGCTCTTCAAAGG 905

Db        781 GGTCTACGTTGCTACTGATTCCGAGAGTTATGTGCCTGAGCTCCAACAGCTCTTCAAAGG 840

Qy        906 GAAGGTGAAGGTGGTGAGCCTGAAGCCTGAGGTGGCCCAGGTCGACCTGTACATCCTCGG 965

Db        841 GAAGGTGAAGGTGGTGAGCCTGAAGCCTGAGGTGGCCCAGGTCGACCTGTACATCCTCGG 900

Qy        966 CCAAGCCGACCACTTTATTGGCAACTGTGTCTCCTCCTTCACTGCCTTTGTGAAGCGGGA 1025

Db        901 CCAAGCCGACCACTTTATTGGCAACTGTGTCTCCTCCTTCACTGCCTTTGTGAAGCGGGA 960

Qy       1026 GCGGGACCTCCAGGGGAGGCCGTCTTCTTTCTTCGGCATGGACAGGCCCCCTAAGCTGCG 1085

Db        961 GCGGGACCTCCAGGGGAGGCCGTCTTCTTTCTTCGGCATGGACAGGCCCCCTAAGCTGCG 1020

Qy       1086 GGACGAGTTCTGATT 1100

Db       1021 GGACGAGTTCTGATT 1035


RESULT 15
BD103191
LOCUS       BD103191               5009 bp    DNA     linear   PAT 27-AUG-2002
DEFINITION  O-fucosyltransferase.
ACCESSION   BD103191
VERSION     BD103191.1  GI:22648765
KEYWORDS    JP 2001527389-A/4.
SOURCE      unidentified
  ORGANISM  unidentified
            unclassified.
REFERENCE   1  (bases 1 to 5009)
  AUTHORS   Wang,Y. and Spellman,M.W.
  TITLE     O-fucosyltransferase
  JOURNAL   Patent: JP 2001527389-A 4 25-DEC-2001;
            GENENTECH INC
COMMENT     OS   Unidentified
            PN   JP 2001527389-A/4
            PD   25-DEC-2001
            PF   17-DEC-1997 JP 1998532877
            PR   31-JAN-1997 US    08/792498,26-NOV-1997 US    08/978741 PI
            YANG WANG,MICHAEL W SPELLMAN
            PC   C12N15/54,C12N9/10,C07K16/40
            CC   Strandedness: Single;
            CC   Topology: Linear;
            CC   O-fucosyltransferase
            FH   Key             Location/Qualifiers
            FT   source          1. .5009
            FT                   /organism='Unidentified'.
FEATURES             Location/Qualifiers
```

```
     source          1. .5009
                     /organism="unidentified"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:32644"
ORIGIN

  Query Match             93.9%;  Score 1033.4;  DB 6;  Length 5009;
  Best Local Similarity   99.9%;  Pred. No. 1e-237;
  Matches 1034;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy         66 GAACCAGGCCGATCACTTCTTGGGCTCTCTGGCATTTGCAAAGCTGCTAAACCGTACCTT 125
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 GAACCAGGCCGATCACTTCTTGGGCTCTCTGGCATTTGCAAAGCTGCTAAACCGTACCTT 60

Qy        126 GGCTGTCCCTCCTTGGATTGAGTACCAGCATCACAAGCCTCCTTTCACCAACCTCCATGT 185
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GGCTGTCCCTCCTTGGATTGAGTACCAGCATCACAAGCCTCCTTTCACCAACCTCCATGT 120

Qy        186 GTCCTACCAGAAGTACTTCAAGCTGGAGCCCCTCCAGGCTTACCATCGGGTCATCAGCTT 245
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 GTCCTACCAGAAGTACTTCAAGCTGGAGCCCCTCCAGGCTTACCATCGGGTCATCAGCTT 180

Qy        246 GGAGGATTTCATGGAGAAGCTGGCACCCACCCACTGGCCCCCTGAGAAGCGGGTGGCATA 305
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 GGAGGATTTCATGGAGAAGCTGGCACCCACCCACTGGCCCCCTGAGAAGCGGGTGGCATA 240

Qy        306 CTGCTTTGAGGTGGCAGCCCAGCGAAGCCCAGATAAGAAGACGTGCCCCATGAAGGAAGG 365
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 CTGCTTTGAGGTGGCAGCCCAGCGAAGCCCAGATAAGAAGACGTGCCCCATGAAGGAAGG 300

Qy        366 AAACCCCTTTGGCCCATTCTGGGATCAGTTTCATGTGAGTTTCAACAAGTCGGAGCTTTT 425
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 AAACCCCTTTGGCCCATTCTGGGATCAGTTTCATGTGAGTTTCAACAAGTCGGAGCTTTT 360

Qy        426 TACAGGCATTTCCTTCAGTGCTTCCTACAGAGAACAATGGAGCCAGAGATTTTCTCCAAA 485
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 TACAGGCATTTCCTTCAGTGCTTCCTACAGAGAACAATGGAGCCAGAGATTTTCTCCAAA 420

Qy        486 GGAACATCCGGTGCTTGCCCTGCCAGGAGCCCCAGCCCAGTTCCCCGTCCTAGAAGAACA 545
              |||||||||||||||||||||||||||||||||||||||||||||||||||| |||||||
Db        421 GGAACATCCGGTGCTTGCCCTGCCAGGAGCCCCAGCCCAGTTCCCCGTCCTAGAGGAACA 480

Qy        546 CAGGCCACTACAGAAGTACATGGTATGGTCAGACGAAATGGTGAAGACGGGAGAGGCCCA 605
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 CAGGCCACTACAGAAGTACATGGTATGGTCAGACGAAATGGTGAAGACGGGAGAGGCCCA 540

Qy        606 GATTCATGCCCACCTTGTCCGGCCCTATGTGGGCATTCATCTGCGCATTGGCTCTGACTG 665
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 GATTCATGCCCACCTTGTCCGGCCCTATGTGGGCATTCATCTGCGCATTGGCTCTGACTG 600

Qy        666 GAAGAACGCCTGTGCCATGCTGAAGGACGGGACTGCAGGCTCGCACTTCATGGCCTCTCC 725
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 GAAGAACGCCTGTGCCATGCTGAAGGACGGGACTGCAGGCTCGCACTTCATGGCCTCTCC 660

Qy        726 GCAGTGTGTGGGCTACAGCCGCAGCACAGCGGCCCCCCTCACGATGACTATGTGCCTGCC 785
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 GCAGTGTGTGGGCTACAGCCGCAGCACAGCGGCCCCCCTCACGATGACTATGTGCCTGCC 720

Qy        786 TGACCTGAAGGAGATCCAGAGGGCTGTGAAGCTCTGGGTGAGGTCGCTGGATGCCCAGTC 845
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 TGACCTGAAGGAGATCCAGAGGGCTGTGAAGCTCTGGGTGAGGTCGCTGGATGCCCAGTC 780

Qy        846 GGTCTACGTTGCTACTGATTCCGAGAGTTATGTGCCTGAGCTCCAACAGCTCTTCAAAGG 905
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 GGTCTACGTTGCTACTGATTCCGAGAGTTATGTGCCTGAGCTCCAACAGCTCTTCAAAGG 840

Qy        906 GAAGGTGAAGGTGGTGAGCCTGAAGCCTGAGGTGGCCCAGGTCGACCTGTACATCCTCGG 965
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 GAAGGTGAAGGTGGTGAGCCTGAAGCCTGAGGTGGCCCAGGTCGACCTGTACATCCTCGG 900

Qy        966 CCAAGCCGACCACTTTATTGGCAACTGTGTCTCCTCCTTCACTGCCTTTGTGAAGCGGGA 1025
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 CCAAGCCGACCACTTTATTGGCAACTGTGTCTCCTCCTTCACTGCCTTTGTGAAGCGGGA 960

Qy       1026 GCGGGACCTCCAGGGGAGGCCGTCTTCTTTCTTCGGCATGGACAGGCCCCCTAAGCTGCG 1085
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 GCGGGACCTCCAGGGGAGGCCGTCTTCTTTCTTCGGCATGGACAGGCCCCCTAAGCTGCG 1020

Qy       1086 GGACGAGTTCTGATT 1100
              |||||||||||||||
Db       1021 GGACGAGTTCTGATT 1035

Search completed: November  9, 2005, 07:59:59
Job time : 4964.62 secs
```

GenCore version 5.1.6

OM nucleic - nucleic search, using sw model

Run on: November 8, 2005, 23:23:13 ; Search time 642.158 Seconds
(without alignments)
10140.369 Million cell updates/sec

Title: US-09-774-954-16
Perfect score: 1100
Sequence: 1 ATGCCCGCGGGCTCCTGGGA.....CTGCGGGACGAGTTCTGATT 1100

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 4390206 seqs, 2959870667 residues

Total number of hits satisfying chosen parameters: 8780412

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_16Dec04:*
1: geneseqn1980s:*
2: geneseqn1990s:*
3: geneseqn2000s:*
4: geneseqn2001as:*
5: geneseqn2001bs:*
6: geneseqn2002as:*
7: geneseqn2002bs:*
8: geneseqn2003as:*
9: geneseqn2003bs:*
10: geneseqn2003cs:*
11: geneseqn2003ds:*
12: geneseqn2004as:*
13: geneseqn2004bs:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	1100	100.0	1514	2	AAV65632	Aav65632 Human hea
2	1098.4	99.9	11284	2	AAV65633	Aav65633 Plasmid c
3	1096.8	99.7	5218	13	ACN40746	Acn40746 Tumour-as
4	1095.8	99.6	5266	10	ADF81754	Adf81754 Leukaemia
5	1046.2	95.1	5230	4	AAK51510	Aak51510 Human pol
6	1041.4	94.7	4850	4	AAK52494	Aak52494 Human pol
7	1033.4	93.9	5009	2	AAV65634	Aav65634 First Eco
8	473.4	43.0	915	4	AAD07629	Aad07629 Human sec
9	406.2	36.9	610	13	ADQ56824	Adq56824 Novel can
10	251.2	22.8	10331	4	AAK79915	Aak79915 Human imm
11	251.2	22.8	10331	4	AAL03344	Aal03344 Human rep
12	233	21.2	479	9	ACH32225	Ach32225 Human end
13	221.8	20.2	477	9	ACH35532	Ach35532 Human end
14	166.4	15.1	1209	4	ABL13947	Abl13947 Drosophil
c 15	150.4	13.7	3264	4	ABL13946	Abl13946 Drosophil
16	123.2	11.2	1545	4	AAF23894	Aaf23894 Human sec
c 17	108.6	9.9	3793	4	ABL13894	Abl13894 Drosophil
c 18	71.8	6.5	3567	4	ABL13952	Abl13952 Drosophil
19	60	5.5	60	6	ABN44539	Abn44539 Human spl
20	56	5.1	760	3	AAZ98210	Aaz98210 Human sig
21	50	4.5	2000	8	ADA71938	Ada71938 Rice gene
c 22	46.2	4.2	2000	8	ADA71938	Ada71938 Rice gene
23	42.4	3.9	1438	8	ACA31561	Aca31561 Prokaryot
c 24	41.2	3.7	114793	4	AAD08215	Aad08215 Human gen
c 25	40.6	3.7	558	5	AAH97782	Aah97782 Murine 7-
26	40.4	3.7	5019	13	ADQ80232	Adq80232 Brain-spe
27	40.4	3.7	5704	12	ADQ84111	Adq84111 Human tum
28	40.4	3.7	5932	10	ADB75488	Adb75488 Prostate
29	40	3.6	40	2	AAV65637	Aav65637 Primer am
30	40	3.6	40	2	AAV65635	Aav65635 Northern
31	40	3.6	1554	8	ADA69815	Ada69815 Rice gene
32	39	3.5	372	8	ACA43606	Aca43606 Prokaryot
33	39	3.5	1131	11	ABD01472	Abd01472 Pseudomon
34	39	3.5	1353	11	ABD01484	Abd01484 Pseudomon
35	38.6	3.5	407	8	ABX46567	Abx46567 Bovine ES
36	38.6	3.5	1596	10	ADI62782	Adi62782 Human apo
37	38.6	3.5	2358	12	ADE77025	Ade77025 Human cDN
38	38.6	3.5	2670	10	ADE25660	Ade25660 Human cDN
39	38.4	3.5	532	9	ACH26305	Ach26305 Human adu
40	38.4	3.5	6227	6	ABT10150	Abt10150 Human bre
41	38.4	3.5	6227	8	ACA64963	Aca64963 Human cor
c 42	37.8	3.4	1234	13	ADS63598	Ads63598 Bacterial
c 43	37.8	3.4	1234	13	ADS63974	Ads63974 Bacterial
c 44	37.8	3.4	1234	13	ADT41577	Adt41577 Bacterial
45	37.8	3.4	2370	4	AAH52044	Aah52044 Mycobacte

ALIGNMENTS

RESULT 1
AAV65632
ID AAV65632 standard; DNA; 1514 BP.
XX
AC AAV65632;
XX
DT 16-DEC-1998 (first entry)
XX
DE Human heart O-fucosyltransferase encoding DNA.
XX
KW O-fucosyltransferase; epidermal growth factor; EGF; glycosylation;
KW O-fucose; inhibitor; sensory neuron; retinal neuron; human; heart; ss.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT misc_feature 1. .1100
FT /*tag= b
FT /note= "this actively expressed O-fucosyltransferase
FT sequence is claimed for in claim 9"
FT CDS 1. .1098
FT /*tag= a
FT /product= "human heart O-fucosyltransferase"
XX
PN WO9833924-A1.
XX
PD 06-AUG-1998.
XX
PF 17-DEC-1997; 97WO-US023401.
XX
PR 31-JAN-1997; 97US-00792498.
PR 26-NOV-1997; 97US-00978741.
XX
PA (GETH) GENENTECH INC.
XX
PI Wang Y, Spellman MW;
XX
DR WPI; 1998-437477/37.
DR P-PSDB; AAW80571.
XX
PT Human O-fucosyl:transferase able to glycosylate epidermal growth factor
PT domains - useful for diagnosis and treatment of diseases involving
PT overexpression of the enzyme.

```
XX
PS   Claim 9; Fig 12A; 90pp; English.
XX
CC   This DNA encodes a human heart O-fucosyltransferase that can glycosylate
CC   an epidermal growth factor (EGF) domain of a polypeptide with an
CC   activated O-fucose residue. Inhibitors of O-fucosyltransferase, e.g.
CC   mutants with increased affinity for the EFG domains, are used in
CC   diagnosis and treatment of conditions associated with overexpression of O
CC   -fucosyltransferase, to promote survival of sensory (retinal) neurons.
CC   Probes based on EGF domain polypeptide are used to detect gene
CC   amplification and expression. The expression can also be determined at
CC   the protein level using antibodies specific for O-fucosyltransferase
XX
SQ   Sequence 1514 BP; 320 A; 435 C; 411 G; 348 T; 0 U; 0 Other;

  Query Match           100.0%;  Score 1100;  DB 2;  Length 1514;
  Best Local Similarity  100.0%;  Pred. No. 1.8e-306;
  Matches 1100;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 ATGCCCGCGGGCTCCTGGGACCCGGCCGGTTACCTGCTCTACTGCCCCTGCATGGGGCGC 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGCCCGCGGGCTCCTGGGACCCGGCCGGTTACCTGCTCTACTGCCCCTGCATGGGGCGC 60

Qy         61 TTTGGGAACCAGGCCGATCACTTCTTGGGCTCTCTGGCATTTGCAAAGCTGCTAAACCGT 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 TTTGGGAACCAGGCCGATCACTTCTTGGGCTCTCTGGCATTTGCAAAGCTGCTAAACCGT 120

Qy        121 ACCTTGGCTGTCCCTCCTTGGATTGAGTACCAGCATCACAAGCCTCCTTTCACCAACCTC 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 ACCTTGGCTGTCCCTCCTTGGATTGAGTACCAGCATCACAAGCCTCCTTTCACCAACCTC 180

Qy        181 CATGTGTCCTACCAGAAGTACTTCAAGCTGGAGCCCCTCCAGGCTTACCATCGGGTCATC 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 CATGTGTCCTACCAGAAGTACTTCAAGCTGGAGCCCCTCCAGGCTTACCATCGGGTCATC 240

Qy        241 AGCTTGGAGGATTTCATGGAGAAGCTGGCACCCACCCACTGGCCCCCTGAGAAGCGGGTG 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 AGCTTGGAGGATTTCATGGAGAAGCTGGCACCCACCCACTGGCCCCCTGAGAAGCGGGTG 300

Qy        301 GCATACTGCTTTGAGGTGGCAGCCCAGCGAAGCCCAGATAAGAAGACGTGCCCCATGAAG 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 GCATACTGCTTTGAGGTGGCAGCCCAGCGAAGCCCAGATAAGAAGACGTGCCCCATGAAG 360

Qy        361 GAAGGAAACCCCTTTGGCCCATTCTGGGATCAGTTTCATGTGAGTTTCAACAAGTCGGAG 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 GAAGGAAACCCCTTTGGCCCATTCTGGGATCAGTTTCATGTGAGTTTCAACAAGTCGGAG 420

Qy        421 CTTTTTACAGGCATTTCCTTCAGTGCTTCCTACAGAGAACAATGGAGCCAGAGATTTTCT 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 CTTTTTACAGGCATTTCCTTCAGTGCTTCCTACAGAGAACAATGGAGCCAGAGATTTTCT 480

Qy        481 CCAAAGGAACATCCGGTGCTTGCCCTGCCAGGAGCCCCAGCCCAGTTCCCCGTCCTAGAA 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 CCAAAGGAACATCCGGTGCTTGCCCTGCCAGGAGCCCCAGCCCAGTTCCCCGTCCTAGAA 540

Qy        541 GAACACAGGCCACTACAGAAGTACATGGTATGGTCAGACGAAATGGTGAAGACGGGAGAG 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 GAACACAGGCCACTACAGAAGTACATGGTATGGTCAGACGAAATGGTGAAGACGGGAGAG 600

Qy        601 GCCCAGATTCATGCCCACCTTGTCCGGCCCTATGTGGGCATTCATCTGCGCATTGGCTCT 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 GCCCAGATTCATGCCCACCTTGTCCGGCCCTATGTGGGCATTCATCTGCGCATTGGCTCT 660

Qy        661 GACTGGAAGAACGCCTGTGCCATGCTGAAGGACGGGACTGCAGGCTCGCACTTCATGGCC 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 GACTGGAAGAACGCCTGTGCCATGCTGAAGGACGGGACTGCAGGCTCGCACTTCATGGCC 720

Qy        721 TCTCCGCAGTGTGTGGGCTACAGCCGCAGCACAGCGGCCCCCCTCACGATGACTATGTGC 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 TCTCCGCAGTGTGTGGGCTACAGCCGCAGCACAGCGGCCCCCCTCACGATGACTATGTGC 780

Qy        781 CTGCCTGACCTGAAGGAGATCCAGAGGGCTGTGAAGCTCTGGGTGAGGTCGCTGGATGCC 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 CTGCCTGACCTGAAGGAGATCCAGAGGGCTGTGAAGCTCTGGGTGAGGTCGCTGGATGCC 840

Qy        841 CAGTCGGTCTACGTTGCTACTGATTCCGAGAGTTATGTGCCTGAGCTCCAACAGCTCTTC 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 CAGTCGGTCTACGTTGCTACTGATTCCGAGAGTTATGTGCCTGAGCTCCAACAGCTCTTC 900

Qy        901 AAAGGGAAGGTGAAGGTGGTGAGCCTGAAGCCTGAGGTGGCCCAGGTCGACCTGTACATC 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 AAAGGGAAGGTGAAGGTGGTGAGCCTGAAGCCTGAGGTGGCCCAGGTCGACCTGTACATC 960

Qy        961 CTCGGCCAAGCCGACCACTTTATTGGCAACTGTGTCTCCTCCTTCACTGCCTTTGTGAAG 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 CTCGGCCAAGCCGACCACTTTATTGGCAACTGTGTCTCCTCCTTCACTGCCTTTGTGAAG 1020

Qy       1021 CGGGAGCGGGACCTCCAGGGGAGGCCGTCTTCTTTCTTCGGCATGGACAGGCCCCCTAAG 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1021 CGGGAGCGGGACCTCCAGGGGAGGCCGTCTTCTTTCTTCGGCATGGACAGGCCCCCTAAG 1080

Qy       1081 CTGCGGGACGAGTTCTGATT 1100
              ||||||||||||||||||||
Db       1081 CTGCGGGACGAGTTCTGATT 1100


RESULT 2
AAV65633
ID   AAV65633 standard; DNA; 11284 BP.
XX
AC   AAV65633;
XX
DT   16-DEC-1998  (first entry)
XX
DE   Plasmid construct for expression of human O-fucosyltransferase.
XX
KW   O-fucosyltransferase; epidermal growth factor; EGF; glycosylation;
KW   O-fucose; inhibitor; sensory neuron; retinal neuron; human; heart; ss.
XX
OS   Homo sapiens.
OS   Synthetic.
XX
FH   Key             Location/Qualifiers
FT   CDS             4140. .5333
FT                   /*tag=  a
FT                   /note= "insert coding for human O-fucosyltransferase."
FT   misc_feature    4218. .4235
FT                   /*tag=  b
FT                   /note= "polyhistidine tag"
XX
PN   WO9833924-A1.
XX
PD   06-AUG-1998.
XX
PF   17-DEC-1997;   97WO-US023401.
XX
PR   31-JAN-1997;   97US-00792498.
PR   26-NOV-1997;   97US-00978741.
XX
PA   (GETH ) GENENTECH INC.
XX
PI   Wang Y,  Spellman MW;
XX
DR   WPI; 1998-437477/37.
DR   P-PSDB; AAW80573.
XX
PT   Human O-fucosyl:transferase able to glycosylate epidermal growth factor
PT   domains - useful for diagnosis and treatment of diseases involving
PT   overexpression of the enzyme.
XX
PS   Example; Page 49-57; 90pp; English.
XX
CC   This represents the nucleotide sequence of the plasmid construct used for
CC   the expression of human heart O-fucosyltransferase. The human O-
CC   fucosyltransferase can glycosylate an epidermal growth factor (EGF)
CC   domain of a polypeptide with an activated O-fucose residue. Inhibitors of
```

```
CC   O-fucosyltransferase, e.g. mutants with increased affinity for the EFG
CC   domains, are used in diagnosis and treatment of conditions associated
CC   with overexpression of O-fucosyltransferase, to promote survival of
CC   sensory (retinal) neurons. Probes based on EGF domain polypeptide are
CC   used to detect gene amplification and expression. The expression can also
CC   be determined at the protein level using antibodies specific for O-
CC   fucosyltransferase
XX
SQ   Sequence 11284 BP; 2947 A; 2602 C; 2624 G; 3111 T; 0 U; 0 Other;

  Query Match             99.9%;  Score 1098.4;  DB 2;  Length 11284;
  Best Local Similarity   99.9%;  Pred. No. 1.3e-305;
  Matches 1099;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 ATGCCCGCGGGCTCCTGGGACCCGGCCGGTTACCTGCTCTACTGCCCCTGCATGGGGCGC 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4236 ATGCCCGCGGGCTCCTGGGACCCGGCCGGTTACCTGCTCTACTGCCCCTGCATGGGGCGC 4295

Qy         61 TTTGGGAACCAGGCCGATCACTTCTTGGGCTCTCTGGCATTTGCAAAGCTGCTAAACCGT 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4296 TTTGGGAACCAGGCCGATCACTTCTTGGGCTCTCTGGCATTTGCAAAGCTGCTAAACCGT 4355

Qy        121 ACCTTGGCTGTCCCTCCTTGGATTGAGTACCAGCATCACAAGCCTCCTTTCACCAACCTC 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4356 ACCTTGGCTGTCCCTCCTTGGATTGAGTACCAGCATCACAAGCCTCCTTTCACCAACCTC 4415

Qy        181 CATGTGTCCTACCAGAAGTACTTCAAGCTGGAGCCCCTCCAGGCTTACCATCGGGTCATC 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4416 CATGTGTCCTACCAGAAGTACTTCAAGCTGGAGCCCCTCCAGGCTTACCATCGGGTCATC 4475

Qy        241 AGCTTGGAGGATTTCATGGAGAAGCTGGCACCCACCCACTGGCCCCCTGAGAAGCGGGTG 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4476 AGCTTGGAGGATTTCATGGAGAAGCTGGCACCCACCCACTGGCCCCCTGAGAAGCGGGTG 4535

Qy        301 GCATACTGCTTTGAGGTGGCAGCCCAGCGAAGCCCAGATAAGAAGACGTGCCCCATGAAG 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4536 GCATACTGCTTTGAGGTGGCAGCCCAGCGAAGCCCAGATAAGAAGACGTGCCCCATGAAG 4595

Qy        361 GAAGGAAACCCCTTTGGCCCATTCTGGGATCAGTTTCATGTGAGTTTCAACAAGTCGGAG 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4596 GAAGGAAACCCCTTTGGCCCATTCTGGGATCAGTTTCATGTGAGTTTCAACAAGTCGGAG 4655

Qy        421 CTTTTTACAGGCATTTCCTTCAGTGCTTCCTACAGAGAACAATGGAGCCAGAGATTTTCT 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4656 CTTTTTACAGGCATTTCCTTCAGTGCTTCCTACAGAGAACAATGGAGCCAGAGATTTTCT 4715

Qy        481 CCAAAGGAACATCCGGTGCTTGCCCTGCCAGGAGCCCCAGCCCAGTTCCCCGTCCTAGAA 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4716 CCAAAGGAACATCCGGTGCTTGCCCTGCCAGGAGCCCCAGCCCAGTTCCCCGTCCTAGAG 4775

Qy        541 GAACACAGGCCACTACAGAAGTACATGGTATGGTCAGACGAAATGGTGAAGACGGGAGAG 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4776 GAACACAGGCCACTACAGAAGTACATGGTATGGTCAGACGAAATGGTGAAGACGGGAGAG 4835

Qy        601 GCCCAGATTCATGCCCACCTTGTCCGGCCCTATGTGGGCATTCATCTGCGCATTGGCTCT 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4836 GCCCAGATTCATGCCCACCTTGTCCGGCCCTATGTGGGCATTCATCTGCGCATTGGCTCT 4895

Qy        661 GACTGGAAGAACGCCTGTGCCATGCTGAAGGACGGGACTGCAGGCTCGCACTTCATGGCC 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4896 GACTGGAAGAACGCCTGTGCCATGCTGAAGGACGGGACTGCAGGCTCGCACTTCATGGCC 4955

Qy        721 TCTCCGCAGTGTGTGGGCTACAGCCGCAGCACAGCGGCCCCCCTCACGATGACTATGTGC 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4956 TCTCCGCAGTGTGTGGGCTACAGCCGCAGCACAGCGGCCCCCCTCACGATGACTATGTGC 5015

Qy        781 CTGCCTGACCTGAAGGAGATCCAGAGGGCTGTGAAGCTCTGGGTGAGGTCGCTGGATGCC 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5016 CTGCCTGACCTGAAGGAGATCCAGAGGGCTGTGAAGCTCTGGGTGAGGTCGCTGGATGCC 5075

Qy        841 CAGTCGGTCTACGTTGCTACTGATTCCGAGAGTTATGTGCCTGAGCTCCAACAGCTCTTC 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5076 CAGTCGGTCTACGTTGCTACTGATTCCGAGAGTTATGTGCCTGAGCTCCAACAGCTCTTC 5135

Qy        901 AAAGGGAAGGTGAAGGTGGTGAGCCTGAAGCCTGAGGTGGCCCAGGTCGACCTGTACATC 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5136 AAAGGGAAGGTGAAGGTGGTGAGCCTGAAGCCTGAGGTGGCCCAGGTCGACCTGTACATC 5195

Qy        961 CTCGGCCAAGCCGACCACTTTATTGGCAACTGTGTCTCCTCCTTCACTGCCTTTGTGAAG 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5196 CTCGGCCAAGCCGACCACTTTATTGGCAACTGTGTCTCCTCCTTCACTGCCTTTGTGAAG 5255

Qy       1021 CGGGAGCGGGACCTCCAGGGGAGGCCGTCTTCTTTCTTCGGCATGGACAGGCCCCCTAAG 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5256 CGGGAGCGGGACCTCCAGGGGAGGCCGTCTTCTTTCTTCGGCATGGACAGGCCCCCTAAG 5315

Qy       1081 CTGCGGGACGAGTTCTGATT 1100
              ||||||||||||||||||||
Db       5316 CTGCGGGACGAGTTCTGATT 5335


RESULT 3
ACN40746
ID   ACN40746 standard; cDNA; 5218 BP.
XX
AC   ACN40746;
XX
DT   18-NOV-2004  (first entry)
XX
DE   Tumour-associated antigenic target (TAT) cDNA DNA326813, SEQ ID NO:5725.
XX
KW   Tumour-associated antigenic target; TAT; human; overexpression; cancer;
KW   tumour; diagnosis; cell proliferative disorder; breast cancer;
KW   colorectal cancer; lung cancer; ovarian cancer; liver cancer;
KW   central nervous system cancer; bladder cancer; pancreatic cancer;
KW   cervical cancer; melanoma; leukaemia; hybridisation probe;
KW   chromosome identification; chromosome mapping; gene mapping;
KW   gene therapy; cytostatic; gene; ss.
XX
OS   Homo sapiens.
XX
PN   WO2004030615-A2.
XX
PD   15-APR-2004.
XX
PF   29-SEP-2003; 2003WO-US028547.
XX
PR   02-OCT-2002; 2002US-0414971P.
XX
PA   (GETH ) GENENTECH INC.
XX
PI   Wu TD,  Zhang Z,  Zhou Y;
XX
DR   WPI; 2004-347921/32.
DR   P-PSDB; ABM82223.
XX
PT   New tumor-associated antigenic target polypeptides and nucleic acids,
PT   useful in preparing a medicament for treating or detecting a
PT   proliferative disorder, e.g. breast, lung, colorectal, ovarian or
PT   prostate cancer or tumor.
XX
PS   Claim 1; SEQ ID NO 5725; 7273pp; English.
XX
CC   The invention relates to human tumour-associated antigenic target (TAT)
CC   polypeptides, and their related nucleic acids. The TAT polypeptides are
CC   overexpressed in cancer tissues compared to normal tissues, and may thus
CC   serve as effective targets for the diagnosis and treatment of cancer in
CC   mammals. The invention also relates to nucleic acid and polypeptide
CC   sequences at least 80% identical to the TAT nucleic acids and
CC   polypeptides; expression vectors and host cells comprising a TAT nucleic
CC   acid; an antibody specific for a TAT polypeptide; a peptide or organic
CC   molecule which binds to a TAT polypeptide; fusion proteins comprising a
CC   TAT polypeptide; and methods and compositions for the treatment or
CC   diagnosis of cancer in mammals. TAT polypeptides, nucleic acids,
CC   antibodies, antagonists, binding molecules and compositions are useful
CC   for diagnosing or treating a cell proliferative disorder associated with
```

```
CC   increased TAT expression, particularly cancers such as breast cancer,
CC   colorectal cancer, lung cancer, ovarian cancer, liver cancer, bladder
CC   cancer, pancreatic cancer, cervical cancer, cancers of the central
CC   nervous system, melanoma and leukaemia. TAT nucleic acids may further be
CC   used as hybridisation probes, in chromosome and gene mapping, in
CC   chromosome identification and in gene therapy. The present sequence
CC   represents a TAT nucleic acid of the invention
XX
SQ   Sequence 5218 BP; 1255 A; 1274 C; 1334 G; 1355 T; 0 U; 0 Other;

  Query Match             99.7%;  Score 1096.8;  DB 13;  Length 5218;
  Best Local Similarity   99.8%;  Pred. No. 2.7e-305;
  Matches 1098;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy          1 ATGCCCGCGGGCTCCTGGGACCCGGCCGGTTACCTGCTCTACTGCCCCTGCATGGGGCGC 60
              ||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        119 ATGCCTGCGGGCTCCTGGGACCCGGCCGGTTACCTGCTCTACTGCCCCTGCATGGGGCGC 178

Qy         61 TTTGGGAACCAGGCCGATCACTTCTTGGGCTCTCTGGCATTTGCAAAGCTGCTAAACCGT 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        179 TTTGGGAACCAGGCCGATCACTTCTTGGGCTCTCTGGCATTTGCAAAGCTGCTAAACCGT 238

Qy        121 ACCTTGGCTGTCCCTCCTTGGATTGAGTACCAGCATCACAAGCCTCCTTTCACCAACCTC 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        239 ACCTTGGCTGTCCCTCCTTGGATTGAGTACCAGCATCACAAGCCTCCTTTCACCAACCTC 298

Qy        181 CATGTGTCCTACCAGAAGTACTTCAAGCTGGAGCCCCTCCAGGCTTACCATCGGGTCATC 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        299 CATGTGTCCTACCAGAAGTACTTCAAGCTGGAGCCCCTCCAGGCTTACCATCGGGTCATC 358

Qy        241 AGCTTGGAGGATTTCATGGAGAAGCTGGCACCCACCCACTGGCCCCCTGAGAAGCGGGTG 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        359 AGCTTGGAGGATTTCATGGAGAAGCTGGCACCCACCCACTGGCCCCCTGAGAAGCGGGTG 418

Qy        301 GCATACTGCTTTGAGGTGGCAGCCCAGCGAAGCCCAGATAAGAAGACGTGCCCCATGAAG 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        419 GCATACTGCTTTGAGGTGGCAGCCCAGCGAAGCCCAGATAAGAAGACGTGCCCCATGAAG 478

Qy        361 GAAGGAAACCCCTTTGGCCCATTCTGGGATCAGTTTCATGTGAGTTTCAACAAGTCGGAG 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        479 GAAGGAAACCCCTTTGGCCCATTCTGGGATCAGTTTCATGTGAGTTTCAACAAGTCGGAG 538

Qy        421 CTTTTTACAGGCATTTCCTTCAGTGCTTCCTACAGAGAACAATGGAGCCAGAGATTTTCT 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        539 CTTTTTACAGGCATTTCCTTCAGTGCTTCCTACAGAGAACAATGGAGCCAGAGATTTTCT 598

Qy        481 CCAAAGGAACATCCGGTGCTTGCCCTGCCAGGAGCCCCAGCCCAGTTCCCCGTCCTAGAA 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        599 CCAAAGGAACATCCGGTGCTTGCCCTGCCAGGAGCCCCAGCCCAGTTCCCCGTCCTAGAG 658

Qy        541 GAACACAGGCCACTACAGAAGTACATGGTATGGTCAGACGAAATGGTGAAGACGGGAGAG 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        659 GAACACAGGCCACTACAGAAGTACATGGTATGGTCAGACGAAATGGTGAAGACGGGAGAG 718

Qy        601 GCCCAGATTCATGCCCACCTTGTCCGGCCCTATGTGGGCATTCATCTGCGCATTGGCTCT 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        719 GCCCAGATTCATGCCCACCTTGTCCGGCCCTATGTGGGCATTCATCTGCGCATTGGCTCT 778

Qy        661 GACTGGAAGAACGCCTGTGCCATGCTGAAGGACGGGACTGCAGGCTCGCACTTCATGGCC 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        779 GACTGGAAGAACGCCTGTGCCATGCTGAAGGACGGGACTGCAGGCTCGCACTTCATGGCC 838

Qy        721 TCTCCGCAGTGTGTGGGCTACAGCCGCAGCACAGCGGCCCCCCTCACGATGACTATGTGC 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        839 TCTCCGCAGTGTGTGGGCTACAGCCGCAGCACAGCGGCCCCCCTCACGATGACTATGTGC 898

Qy        781 CTGCCTGACCTGAAGGAGATCCAGAGGGCTGTGAAGCTCTGGGTGAGGTCGCTGGATGCC 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        899 CTGCCTGACCTGAAGGAGATCCAGAGGGCTGTGAAGCTCTGGGTGAGGTCGCTGGATGCC 958

Qy        841 CAGTCGGTCTACGTTGCTACTGATTCCGAGAGTTATGTGCCTGAGCTCCAACAGCTCTTC 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        959 CAGTCGGTCTACGTTGCTACTGATTCCGAGAGTTATGTGCCTGAGCTCCAACAGCTCTTC 1018

Qy        901 AAAGGGAAGGTGAAGGTGGTGAGCCTGAAGCCTGAGGTGGCCCAGGTCGACCTGTACATC 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1019 AAAGGGAAGGTGAAGGTGGTGAGCCTGAAGCCTGAGGTGGCCCAGGTCGACCTGTACATC 1078

Qy        961 CTCGGCCAAGCCGACCACTTTATTGGCAACTGTGTCTCCTCCTTCACTGCCTTTGTGAAG 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1079 CTCGGCCAAGCCGACCACTTTATTGGCAACTGTGTCTCCTCCTTCACTGCCTTTGTGAAG 1138

Qy       1021 CGGGAGCGGGACCTCCAGGGGAGGCCGTCTTCTTTCTTCGGCATGGACAGGCCCCCTAAG 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1139 CGGGAGCGGGACCTCCAGGGGAGGCCGTCTTCTTTCTTCGGCATGGACAGGCCCCCTAAG 1198

Qy       1081 CTGCGGGACGAGTTCTGATT 1100
              ||||||||||||||||||||
Db       1199 CTGCGGGACGAGTTCTGATT 1218


RESULT 4
ADF81754
ID   ADF81754 standard; DNA; 5266 BP.
XX
AC   ADF81754;
XX
DT   26-FEB-2004  (first entry)
XX
DE   Leukaemia-related DNA sequence #2310.
XX
KW   Cytostatic; Gene therapy; leukaemia; ss.
XX
OS   Unidentified.
XX
PN   WO2003039443-A2.
XX
PD   15-MAY-2003.
XX
PF   04-NOV-2002; 2002WO-EP012303.
XX
PR   05-NOV-2001; 2001EP-00126244.
PR   30-APR-2002; 2002EP-00009758.
XX
PA   (DEKR-) DEUT KREBSFORSCHUNGSZENTRUM.
PA   (UYLU-) UNIV LUDWIG MAXIMILIANS.
PA   (HAFE/) HAFERLACH T.
PA   (SCHO/) SCHOCH C.
PA   (KERN/) KERN W.
XX
PI   Haferlach T,  Schoch C,  Kern W,  Kohlmann A,  Schnittger S,  Dugas M;
PI   Eils R,  Brors B,  Mergenthaler S;
XX
DR   WPI; 2003-505037/47.
XX
PT   Determining the subtype of leukemia cells and whether a patient sample
PT   contains leukemia cells or other cells, useful for treating leukemia,
PT   comprises determining the expression profile of a group of markers in a
PT   patient sample.
XX
PS   Disclosure; SEQ ID NO 2310; 2938pp; English.
XX
CC   The present invention relates to a method (M1) for determining the
CC   subtype of leukaemia cells and whether a patient sample contains
CC   leukaemia cells. The method comprises determining the expression profile
CC   of a group of markers in a patient sample. The method is useful for
CC   determining the presence of leukaemia cells, its types or subtypes, and
CC   for the preparation of a medicament for treating leukaemia.
XX
SQ   Sequence 5266 BP; 1280 A; 1278 C; 1331 G; 1346 T; 0 U; 31 Other;

  Query Match             99.6%;  Score 1095.8;  DB 10;  Length 5266;
  Best Local Similarity   99.7%;  Pred. No. 5.2e-305;
  Matches 1097;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy          1 ATGCCCGCGGGCTCCTGGGACCCGGCCGGTTACCTGCTCTACTGCCCCTGCATGGGGCGC 60
```

```
Db        132 ATGCCTGCGGGCTCCTGGGACCCGGCCGGTTACCTGCTCTACTGCCCCTGCATGGGGCGC 191

Qy         61 TTTGGGAACCAGGCCGATCACTTCTTGGGCTCTCTGGCATTTGCAAAGCTGCTAAACCGT 120

Db        192 TTTGGGAACCAGGCCGATCACTTCTTGGGCTCTCTGGCATTTGCAAAGCTGCTAAACCGT 251

Qy        121 ACCTTGGCTGTCCCTCCTTGGATTGAGTACCAGCATCACAAGCCTCCTTTCACCAACCTC 180

Db        252 ACCTTGGCTGTCCCTCCTTGGATTGAGTACCAGCATCACAAGCCTCCTTTCACCAACCTC 311

Qy        181 CATGTGTCCTACCAGAAGTACTTCAAGCTGGAGCCCCTCCAGGCTTACCATCGGGTCATC 240

Db        312 CATGTGTCCTACCAGAAGTACTTCAAGCTGGAGCCCCTCCAGGCTTACCATCGGGTCATC 371

Qy        241 AGCTTGGAGGATTTCATGGAGAAGCTGGCACCCACCCACTGGCCCCCTGAGAAGCGGGTG 300

Db        372 AGCTTGGAGGATTTCATGGAGAAGCTGGCACCCACCCACTGGCCCCCTGAGAAGCGGGTG 431

Qy        301 GCATACTGCTTTGAGGTGGCAGCCCAGCGAAGCCCAGATAAGAAGACGTGCCCCATGAAG 360

Db        432 GCATACTGCTTTGAGGTGGCAGCCCAGCGAAGCCCAGATAAGAAGACGTGCCCCATGAAG 491

Qy        361 GAAGGAAACCCCTTTGGCCCATTCTGGGATCAGTTTCATGTGAGTTTCAACAAGTCGGAG 420

Db        492 GAAGGAAACCCCTTTGGCCCATTCTGGGATCAGTTTCATGTGAGTTTCAACAAGTCGGAG 551

Qy        421 CTTTTTACAGGCATTTCCTTCAGTGCTTCCTACAGAGAACAATGGAGCCAGAGATTTTCT 480

Db        552 CTTTTTACAGGCATTTCCTTCAGTGCTTCCTACAGAGAACAATGGAGCCAGAGATTTTCT 611

Qy        481 CCAAAGGAACATCCGGTGCTTGCCCTGCCAGGAGCCCCAGCCCAGTTCCCCGTCCTAGAA 540

Db        612 CCAAAGGAACATCCGGTGCTTGCCCTGCCAGGAGCCCCAGCCCAGTTCCCCGTCCTAGAG 671

Qy        541 GAACACAGGCCACTACAGAAGTACATGGTATGGTCAGACGAAATGGTGAAGACGGGAGAG 600

Db        672 GAACACAGGCCACTACAGAAGTACATGGTATGGTCAGACGAAATGGTGAAGACGGGAGAG 731

Qy        601 GCCCAGATTCATGCCCACCTTGTCCGGCCCTATGTGGGCATTCATCTGCGCATTGGCTCT 660

Db        732 GCCCAGATTCATGCCCACCTTGTCCGGCCCTATGTGGGCATTCATCTGCGCATTGGCTCT 791

Qy        661 GACTGGAAGAACGCCTGTGCCATGCTGAAGGACGGGACTGCAGGCTCGCACTTCATGGCC 720

Db        792 GACTGGAAGAACGCCTGTGCCATGCTGAAGGACGGGACTGCAGGCTCGCACTTCATGGCC 851

Qy        721 TCTCCGCAGTGTGTGGGCTACAGCCGCAGCACAGCGGCCCCCCTCACGATGACTATGTGC 780

Db        852 TCTCCGCAGTGTGTGGGCTACAGCCGCAGCACAGCGGCCCCCCTCACGATGACTATGTGC 911

Qy        781 CTGCCTGACCTGAAGGAGATCCAGAGGGCTGTGAAGCTCTGGGTGAGGTCGCTGGATGCC 840

Db        912 CTGCCTGACCTGAAGGAGATCCAGAGGGCTGTGAAGCTCTGGGTGAGGTCGCTGGATGCC 971

Qy        841 CAGTCGGTCTACGTTGCTACTGATTCCGAGAGTTATGTGCCTGAGCTCCAACAGCTCTTC 900

Db        972 CAGTCGGTCTACGTTGCTACTGATTCCGAGAGTTATGTGCCTGAGCTCCAACAGCTCTTC 1031

Qy        901 AAAGGGAAGGTGAAGGTGGTGAGCCTGAAGCCTGAGGTGGCCCAGGTCGACCTGTACATC 960

Db       1032 AAAGGGAAGGTGAAGGTGGTGAGCCTGAAGCCTGAGGTGGCCCAGGTCGACCTGTACATC 1091

Qy        961 CTCGGCCAAGCCGACCACTTTATTGGCAACTGTGTCTCCTCCTTCACTGCCTTTGTGAAG 1020

Db       1092 CTCGGNCAAGCCGACCACTTTATTGGCAACTGTGTCTCCTCCTTCACTGCCTTTGTGAAG 1151

Qy       1021 CGGGAGCGGGACCTCCAGGGGAGGCCGTCTTCTTTCTTCGGCATGGACAGGCCCCCTAAG 1080

Db       1152 CGGGAGCGGGACCTCCAGGGGAGGCCGTCTTCTTTCTTCGGCATGGACAGGCCCCCTAAG 1211

Qy       1081 CTGCGGGACGAGTTCTGATT 1100

Db       1212 CTGCGGGACGAGTTCTGATT 1231
```

RESULT 5
AAK51510
ID AAK51510 standard; cDNA; 5230 BP.
XX
AC AAK51510;
XX
DT 06-NOV-2001 (first entry)
XX
DE Human polynucleotide SEQ ID NO 55.
XX
KW Human; cytokine; cell proliferation; cell differentiation; gene therapy;
KW vaccine; peptide therapy; stem cell growth factor; haematopoiesis;
KW tissue growth factor; immunomodulatory; cancer; leukaemia;
KW nervous system disorder; arthritis; inflammation; ss.
XX
OS Homo sapiens.
XX
PN WO200157190-A2.
XX
PD 09-AUG-2001.
XX
PF 05-FEB-2001; 2001WO-US004098.
XX
PR 03-FEB-2000; 2000US-00496914.
PR 27-APR-2000; 2000US-00560875.
PR 20-JUN-2000; 2000US-00598075.
PR 19-JUL-2000; 2000US-00620325.
PR 01-SEP-2000; 2000US-00654936.
PR 15-SEP-2000; 2000US-00663561.
PR 20-OCT-2000; 2000US-00693325.
PR 30-NOV-2000; 2000US-00728422.
XX
PA (HYSE-) HYSEQ INC.
XX
PI Tang YT, Liu C, Drmanac RT, Asundi V, Zhou P, Xu C, Cao Y;
PI Ma Y, Zhao QA, Wang D, Wang J, Zhang J, Ren F, Chen R, Wang ZW;
PI Xue AJ, Yang Y, Wejhrman T, Goodrich R;
XX
DR WPI; 2001-476283/51.
DR P-PSDB; AAM78377.
XX
PT Nucleic acids encoding polypeptides with cytokine-like activities, useful
PT in diagnosis and gene therapy.
XX
PS Claim 1; Page 638-642; 6221pp; English.
XX
CC The invention relates to polynucleotides (AAK51456-AAK53435) and the
CC encoded polypeptides (AAM78323-AAM80302) that exhibit activity elating to
CC cytokine, cell proliferation or cell differentiation or which may induce
CC production of other cytokines in other cell populations. The
CC polynucleotides and polypeptides are useful in gene therapy, vaccines or
CC peptide therapy. The polypeptides have various cytokine-like activities,
CC e.g. stem cell growth factor activity, haematopoiesis regulating
CC activity, tissue growth factor activity, immunomodulatory activity and
CC activin/inhibin activity and may be useful in the diagnosis and/or
CC treatment of cancer, leukaemia, nervous system disorders, arthritis and
CC inflammation. Note: Records for SEQ ID NO 2110 (AAK52581), 2111
CC (AAK52582) and 3666 (AAM80020) are omitted as the relevant pages from the
CC sequence listing were missing at the time of publication
XX
SQ Sequence 5230 BP; 1285 A; 1255 C; 1330 G; 1360 T; 0 U; 0 Other;

Query Match 95.1%; Score 1046.2; DB 4; Length 5230;
Best Local Similarity 99.7%; Pred. No. 1.1e-290;
Matches 1048; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

```
Qy         50 GCATGGGGCGCTTTGGGAACCAGGCCGATCACTTCTTGGGCTCTCTGGCATTTGCAAAGC 109

Db        206 GCCAGGGGCGCTTTGGGAACCAGGCCGATCACTTCTTGGGCTCTCTGGCATTTGCAAAGC 265
```

```
Qy        110 TGCTAAACCGTACCTTGGCTGTCCCTCCTTGGATTGAGTACCAGCATCACAAGCCTCCTT 169
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        266 TGCTAAACCGTACCTTGGCTGTCCCTCCTTGGATTGAGTACCAGCATCACAAGCCTCCTT 325

Qy        170 TCACCAACCTCCATGTGTCCTACCAGAAGTACTTCAAGCTGGAGCCCCTCCAGGCTTACC 229
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        326 TCACCAACCTCCATGTGTCCTACCAGAAGTACTTCAAGCTGGAGCCCCTCCAGGCTTACC 385

Qy        230 ATCGGGTCATCAGCTTGGAGGATTTCATGGAGAAGCTGGCACCCACCCACTGGCCCCCTG 289
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        386 ATCGGGTCATCAGCTTGGAGGATTTCATGGAGAAGCTGGCACCCACCCACTGGCCCCCTG 445

Qy        290 AGAAGCGGGTGGCATACTGCTTTGAGGTGGCAGCCCAGCGAAGCCCAGATAAGAAGACGT 349
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        446 AGAAGCGGGTGGCATACTGCTTTGAGGTGGCAGCCCAGCGAAGCCCAGATAAGAAGACGT 505

Qy        350 GCCCCATGAAGGAAGGAAACCCCTTTGGCCCATTCTGGGATCAGTTTCATGTGAGTTTCA 409
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        506 GCCCCATGAAGGAAGGAAACCCCTTTGGCCCATTCTGGGATCAGTTTCATGTGAGTTTCA 565

Qy        410 ACAAGTCGGAGCTTTTTACAGGCATTTCCTTCAGTGCTTCCTACAGAGAACAATGGAGCC 469
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        566 ACAAGTCGGAGCTTTTTACAGGCATTTCCTTCAGTGCTTCCTACAGAGAACAATGGAGCC 625

Qy        470 AGAGATTTTCTCCAAAGGAACATCCGGTGCTTGCCCTGCCAGGAGCCCCAGCCCAGTTCC 529
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        626 AGAGATTTTCTCCAAAGGAACATCCGGTGCTTGCCCTGCCAGGAGCCCCAGCCCAGTTCC 685

Qy        530 CCGTCCTAGAAGAACACAGGCCACTACAGAAGTACATGGTATGGTCAGACGAAATGGTGA 589
              ||||||||| ||||||||||||||||||||||||||||||||||||||||||||||||||
Db        686 CCGTCCTAGAGGAACACAGGCCACTACAGAAGTACATGGTATGGTCAGACGAAATGGTGA 745

Qy        590 AGACGGGAGAGGCCCAGATTCATGCCCACCTTGTCCGGCCCTATGTGGGCATTCATCTGC 649
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        746 AGACGGGAGAGGCCCAGATTCATGCCCACCTTGTCCGGCCCTATGTGGGCATTCATCTGC 805

Qy        650 GCATTGGCTCTGACTGGAAGAACGCCTGTGCCATGCTGAAGGACGGGACTGCAGGCTCGC 709
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        806 GCATTGGCTCTGACTGGAAGAACGCCTGTGCCATGCTGAAGGACGGGACTGCAGGCTCGC 865

Qy        710 ACTTCATGGCCTCTCCGCAGTGTGTGGGCTACAGCCGCAGCACAGCGGCCCCCCTCACGA 769
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        866 ACTTCATGGCCTCTCCGCAGTGTGTGGGCTACAGCCGCAGCACAGCGGCCCCCCTCACGA 925

Qy        770 TGACTATGTGCCTGCCTGACCTGAAGGAGATCCAGAGGGCTGTGAAGCTCTGGGTGAGGT 829
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        926 TGACTATGTGCCTGCCTGACCTGAAGGAGATCCAGAGGGCTGTGAAGCTCTGGGTGAGGT 985

Qy        830 CGCTGGATGCCCAGTCGGTCTACGTTGCTACTGATTCCGAGAGTTATGTGCCTGAGCTCC 889
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        986 CGCTGGATGCCCAGTCGGTCTACGTTGCTACTGATTCCGAGAGTTATGTGCCTGAGCTCC 1045

Qy        890 AACAGCTCTTCAAAGGGAAGGTGAAGGTGGTGAGCCTGAAGCCTGAGGTGGCCCAGGTCG 949
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1046 AACAGCTCTTCAAAGGGAAGGTGAAGGTGGTGAGCCTGAAGCCTGAGGTGGCCCAGGTCG 1105

Qy        950 ACCTGTACATCCTCGGCCAAGCCGACCACTTTATTGGCAACTGTGTCTCCTCCTTCACTG 1009
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1106 ACCTGTACATCCTCGGCCAAGCCGACCACTTTATTGGCAACTGTGTCTCCTCCTTCACTG 1165

Qy       1010 CCTTTGTGAAGCGGGAGCGGGACCTCCAGGGGAGGCCGTCTTCTTTCTTCGGCATGGACA 1069
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1166 CCTTTGTGAAGCGGGAGCGGGACCTCCAGGGGAGGCCGTCTTCTTTCTTCGGCATGGACA 1225

Qy       1070 GGCCCCCTAAGCTGCGGGACGAGTTCTGATT 1100
              |||||||||||||||||||||||||||||||
Db       1226 GGCCCCCTAAGCTGCGGGACGAGTTCTGATT 1256
```

RESULT 6
AAK52494
ID AAK52494 standard; cDNA; 4850 BP.
XX

AC AAK52494;
XX
DT 06-NOV-2001 (first entry)
XX
DE Human polynucleotide SEQ ID NO 2023.
XX
KW Human; cytokine; cell proliferation; cell differentiation; gene therapy;
KW vaccine; peptide therapy; stem cell growth factor; haematopoiesis;
KW tissue growth factor; immunomodulatory; cancer; leukaemia;
KW nervous system disorder; arthritis; inflammation; ss.
XX
OS Homo sapiens.
XX
PN WO200157190-A2.
XX
PD 09-AUG-2001.
XX
PF 05-FEB-2001; 2001WO-US004098.
XX
PR 03-FEB-2000; 2000US-00496914.
PR 27-APR-2000; 2000US-00560875.
PR 20-JUN-2000; 2000US-00598075.
PR 19-JUL-2000; 2000US-00620325.
PR 01-SEP-2000; 2000US-00654936.
PR 15-SEP-2000; 2000US-00663561.
PR 20-OCT-2000; 2000US-00693325.
PR 30-NOV-2000; 2000US-00728422.
XX
PA (HYSE-) HYSEQ INC.
XX
PI Tang YT, Liu C, Drmanac RT, Asundi V, Zhou P, Xu C, Cao Y;
PI Ma Y, Zhao QA, Wang D, Wang J, Zhang J, Ren F, Chen R, Wang ZW;
PI Xue AJ, Yang Y, Wejhrman T, Goodrich R;
XX
DR WPI; 2001-476283/51.
DR P-PSDB; AAM79361.
XX
PT Nucleic acids encoding polypeptides with cytokine-like activities, useful
PT in diagnosis and gene therapy.
XX
PS Claim 1; Page 4424-4425; 6221pp; English.
XX
CC The invention relates to polynucleotides (AAK51456-AAK53435) and the
CC encoded polypeptides (AAM78323-AAM80302) that exhibit activity elating to
CC cytokine, cell proliferation or cell differentiation or which may induce
CC production of other cytokines in other cell populations. The
CC polynucleotides and polypeptides are useful in gene therapy, vaccines or
CC peptide therapy. The polypeptides have various cytokine-like activities,
CC e.g. stem cell growth factor activity, haematopoiesis regulating
CC activity, tissue growth factor activity, immunomodulatory activity and
CC activin/inhibin activity and may be useful in the diagnosis and/or
CC treatment of cancer, leukaemia, nervous system disorders, arthritis and
CC inflammation. Note: Records for SEQ ID NO 2110 (AAK52581), 2111
CC (AAK52582) and 3666 (AAM80020) are omitted as the relevant pages from the
CC sequence listing were missing at the time of publication
XX
SQ Sequence 4850 BP; 1142 A; 1218 C; 1244 G; 1246 T; 0 U; 0 Other;

```
  Query Match             94.7%;  Score 1041.4;  DB 4;  Length 4850;
  Best Local Similarity   99.4%;  Pred. No. 2.5e-289;
  Matches 1045;  Conservative    0;  Mismatches    6;  Indels    0;  Gaps    0;

Qy          50 GCATGGGGCGCTTTGGGAACCAGGCCGATCACTTCTTGGGCTCTCTGGCATTTGCAAAGC 109
               ||  ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         206 GCCAGGGGCGCTTTGGGAACCAGGCCGATCACTTCTTGGGCTCTCTGGCATTTGCAAAGC 265

Qy         110 TGCTAAACCGTACCTTGGCTGTCCCTCCTTGGATTGAGTACCAGCATCACAAGCCTCCTT 169
               |||| |||||| |||||||||||||| ||||||||||||||||||||||||||||||||
Db         266 TGCTGAACCGTTCCTTGGCTGTCCCTTCTTGGATTGAGTACCAGCATCACAAGCCTCCTT 325

Qy         170 TCACCAACCTCCATGTGTCCTACCAGAAGTACTTCAAGCTGGAGCCCCTCCAGGCTTACC 229
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         326 TCACCAACCTCCATGTGTCCTACCAGAAGTACTTCAAGCTGGAGCCCCTCCAGGCTTACC 385
```

```
Qy        230 ATCGGGTCATCAGCTTGGAGGATTTCATGGAGAAGCTGGCACCCACCCACTGGCCCCCTG 289
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        386 ATCGGGTCATCAGCTTGGAGGATTTCATGGAGAAGCTGGCACCCACCCACTGGCCCCCTG 445

Qy        290 AGAAGCGGGTGGCATACTGCTTTGAGGTGGCAGCCCAGCGAAGCCCAGATAAGAAGACGT 349
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        446 AGAAGCGGGTGGCATACTGCTTTGAGGTGGCAGCCCAGCGAAGCCCAGATAAGAAGACGT 505

Qy        350 GCCCCATGAAGGAAGGAAACCCCTTTGGCCCATTCTGGGATCAGTTTCATGTGAGTTTCA 409
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        506 GCCCCATGAAGGAAGGAAACCCCTTTGGCCCATTCTGGGATCAGTTTCATGTGAGTTTCA 565

Qy        410 ACAAGTCGGAGCTTTTTACAGGCATTTCCTTCAGTGCTTCCTACAGAGAACAATGGAGCC 469
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        566 ACAAGTCGGAGCTTTTTACAGGCATTTCCTTCAGTGCTTCCTACAGAGAACAATGGAGCC 625

Qy        470 AGAGATTTTCTCCAAAGGAACATCCGGTGCTTGCCCTGCCAGGAGCCCCAGCCCAGTTCC 529
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        626 AGAGATTTTCTCCAAAGGAACATCCGGTGCTTGCCCTGCCAGGAGCCCCAGCCCAGTTCC 685

Qy        530 CCGTCCTAGAAGAACACAGGCCACTACAGAAGTACATGGTATGGTCAGACGAAATGGTGA 589
              |||||||||| ||||||||||||||||||||||||||||||||||||||||||||||||||
Db        686 CCGTCCTAGAGGAACACAGGCCACTACAGAAGTACATGGTATGGTCAGACGAAATGGTGA 745

Qy        590 AGACGGGAGAGGCCCAGATTCATGCCCACCTTGTCCGGCCCTATGTGGGCATTCATCTGC 649
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        746 AGACGGGAGAGGCCCAGATTCATGCCCACCTTGTCCGGCCCTATGTGGGCATTCATCTGC 805

Qy        650 GCATTGGCTCTGACTGGAAGAACGCCTGTGCCATGCTGAAGGACGGGACTGCAGGCTCGC 709
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        806 GCATTGGCTCTGACTGGAAGAACGCCTGTGCCATGCTGAAGGACGGGACTGCAGGCTCGC 865

Qy        710 ACTTCATGGCCTCTCCGCAGTGTGTGGGCTACAGCCGCAGCACAGCGGCCCCCCTCACGA 769
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        866 ACTTCATGGCCTCTCCGCAGTGTGTGGGCTACAGCCGCAGCACAGCGGCCCCCCTCACGA 925

Qy        770 TGACTATGTGCCTGCCTGACCTGAAGGAGATCCAGAGGGCTGTGAAGCTCTGGGTGAGGT 829
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        926 TGACTATGTGCCTGCCTGACCTGAAGGAGATCCAGAGGGCTGTGAAGCTCTGGGTGAGGT 985

Qy        830 CGCTGGATGCCCAGTCGGTCTACGTTGCTACTGATTCCGAGAGTTATGTGCCTGAGCTCC 889
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        986 CGCTGGATGCCCAGTCGGTCTACGTTGCTACTGATTCCGAGAGTTATGTGCCTGAGCTCC 1045

Qy        890 AACAGCTCTTCAAAGGGAAGGTGAAGGTGGTGAGCCTGAAGCCTGAGGTGGCCCAGGTCG 949
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1046 AACAGCTCTTCAAAGGGAAGGTGAAGGTGGTGAGCCTGAAGCCTGAGGTGGCCCAGGTCG 1105

Qy        950 ACCTGTACATCCTCGGCCAAGCCGACCACTTTATTGGCAACTGTGTCTCCTCCTTCACTG 1009
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1106 ACCTGTACATCCTCGGCCAAGCCGACCACTTTATTGGCAACTGTGTCTCCTCCTTCACTG 1165

Qy       1010 CCTTTGTGAAGCGGGAGCGGGACCTCCAGGGGAGGCCGTCTTCTTTCTTCGGCATGGACA 1069
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1166 CCTTTGTGAAGCGGGAGCGGGACCTCCAGGGGAGGCCGTCTTCTTTCTTCGGCATGGACA 1225

Qy       1070 GGCCCCCTAAGCTGCGGGACGAGTTCTGATT 1100
              |||||||||||||||||||||||||||||||
Db       1226 GGCCCCCTAAGCTGCGGGACGAGTTCTGATT 1256
```

RESULT 7
AAV65634
ID AAV65634 standard; cDNA; 5009 BP.
XX
AC AAV65634;
XX
DT 16-DEC-1998 (first entry)
XX
DE First EcoR1 nucleotide fragment of human KIAA0180.
XX
KW O-fucosyltransferase; epidermal growth factor; EGF; glycosylation;
KW O-fucose; inhibitor; sensory neuron; retinal neuron; human; heart; ss.
XX
OS Homo sapiens.
XX
PN WO9833924-A1.
XX
PD 06-AUG-1998.
XX
PF 17-DEC-1997; 97WO-US023401.
XX
PR 31-JAN-1997; 97US-00792498.
PR 26-NOV-1997; 97US-00978741.
XX
PA (GETH) GENENTECH INC.
XX
PI Wang Y, Spellman MW;
XX
DR WPI; 1998-437477/37.
DR P-PSDB; AAW80577.
XX
PT Human O-fucosyl:transferase able to glycosylate epidermal growth factor
PT domains - useful for diagnosis and treatment of diseases involving
PT overexpression of the enzyme.
XX
PS Example 1; Fig 11; 90pp; English.
XX
CC This represents a first EcoR1 nucleotide fragment of human KIAA0180. This
CC 5009 basepairs partial cDNA encodes for a protein of unknown function
CC from myeblast cel line KG-1. The invention provides a human heart O-
CC fucosyltransferase that can glycosylate an epidermal growth factor (EGF)
CC domain of a polypeptide with an activated O-fucose residue. Inhibitors of
CC O-fucosyltransferase, e.g. mutants with increased affinity for the EFG
CC domains, are used in diagnosis and treatment of conditions associated
CC with overexpression of O-fucosyltransferase, to promote survival of
CC sensory (retinal) neurons. Probes based on EGF domain polypeptide are
CC used to detect gene amplification and expression. The expression can also
CC be determined at the protein level using antibodies specific for O-
CC fucosyltransferase
XX
SQ Sequence 5009 BP; 1235 A; 1195 C; 1268 G; 1311 T; 0 U; 0 Other;

Query Match 93.9%; Score 1033.4; DB 2; Length 5009;
Best Local Similarity 99.9%; Pred. No. 5.1e-287;
Matches 1034; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```
Qy         66 GAACCAGGCCGATCACTTCTTGGGCTCTCTGGCATTTGCAAAGCTGCTAAACCGTACCTT 125
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 GAACCAGGCCGATCACTTCTTGGGCTCTCTGGCATTTGCAAAGCTGCTAAACCGTACCTT 60

Qy        126 GGCTGTCCCTCCTTGGATTGAGTACCAGCATCACAAGCCTCCTTTCACCAACCTCCATGT 185
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GGCTGTCCCTCCTTGGATTGAGTACCAGCATCACAAGCCTCCTTTCACCAACCTCCATGT 120

Qy        186 GTCCTACCAGAAGTACTTCAAGCTGGAGCCCCTCCAGGCTTACCATCGGGTCATCAGCTT 245
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 GTCCTACCAGAAGTACTTCAAGCTGGAGCCCCTCCAGGCTTACCATCGGGTCATCAGCTT 180

Qy        246 GGAGGATTTCATGGAGAAGCTGGCACCCACCCACTGGCCCCCTGAGAAGCGGGTGGCATA 305
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 GGAGGATTTCATGGAGAAGCTGGCACCCACCCACTGGCCCCCTGAGAAGCGGGTGGCATA 240

Qy        306 CTGCTTTGAGGTGGCAGCCCAGCGAAGCCCAGATAAGAAGACGTGCCCCATGAAGGAAGG 365
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 CTGCTTTGAGGTGGCAGCCCAGCGAAGCCCAGATAAGAAGACGTGCCCCATGAAGGAAGG 300

Qy        366 AAACCCCTTTGGCCCATTCTGGGATCAGTTTCATGTGAGTTTCAACAAGTCGGAGCTTTT 425
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 AAACCCCTTTGGCCCATTCTGGGATCAGTTTCATGTGAGTTTCAACAAGTCGGAGCTTTT 360

Qy        426 TACAGGCATTTCCTTCAGTGCTTCCTACAGAGAACAATGGAGCCAGAGATTTTCTCCAAA 485
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 TACAGGCATTTCCTTCAGTGCTTCCTACAGAGAACAATGGAGCCAGAGATTTTCTCCAAA 420
```

```
Qy        486 GGAACATCCGGTGCTTGCCCTGCCAGGAGCCCCAGCCCAGTTCCCCGTCCTAGAAGAACA 545
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||| |||||
Db        421 GGAACATCCGGTGCTTGCCCTGCCAGGAGCCCCAGCCCAGTTCCCCGTCCTAGAGGAACA 480

Qy        546 CAGGCCACTACAGAAGTACATGGTATGGTCAGACGAAATGGTGAAGACGGGAGAGGCCCA 605
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 CAGGCCACTACAGAAGTACATGGTATGGTCAGACGAAATGGTGAAGACGGGAGAGGCCCA 540

Qy        606 GATTCATGCCCACCTTGTCCGGCCCTATGTGGGCATTCATCTGCGCATTGGCTCTGACTG 665
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 GATTCATGCCCACCTTGTCCGGCCCTATGTGGGCATTCATCTGCGCATTGGCTCTGACTG 600

Qy        666 GAAGAACGCCTGTGCCATGCTGAAGGACGGGACTGCAGGCTCGCACTTCATGGCCTCTCC 725
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 GAAGAACGCCTGTGCCATGCTGAAGGACGGGACTGCAGGCTCGCACTTCATGGCCTCTCC 660

Qy        726 GCAGTGTGTGGGCTACAGCCGCAGCACAGCGGCCCCCCTCACGATGACTATGTGCCTGCC 785
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 GCAGTGTGTGGGCTACAGCCGCAGCACAGCGGCCCCCCTCACGATGACTATGTGCCTGCC 720

Qy        786 TGACCTGAAGGAGATCCAGAGGGCTGTGAAGCTCTGGGTGAGGTCGCTGGATGCCCAGTC 845
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 TGACCTGAAGGAGATCCAGAGGGCTGTGAAGCTCTGGGTGAGGTCGCTGGATGCCCAGTC 780

Qy        846 GGTCTACGTTGCTACTGATTCCGAGAGTTATGTGCCTGAGCTCCAACAGCTCTTCAAAGG 905
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 GGTCTACGTTGCTACTGATTCCGAGAGTTATGTGCCTGAGCTCCAACAGCTCTTCAAAGG 840

Qy        906 GAAGGTGAAGGTGGTGAGCCTGAAGCCTGAGGTGGCCCAGGTCGACCTGTACATCCTCGG 965
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 GAAGGTGAAGGTGGTGAGCCTGAAGCCTGAGGTGGCCCAGGTCGACCTGTACATCCTCGG 900

Qy        966 CCAAGCCGACCACTTTATTGGCAACTGTGTCTCCTCCTTCACTGCCTTTGTGAAGCGGGA 1025
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 CCAAGCCGACCACTTTATTGGCAACTGTGTCTCCTCCTTCACTGCCTTTGTGAAGCGGGA 960

Qy       1026 GCGGGACCTCCAGGGGAGGCCGTCTTCTTTCTTCGGCATGGACAGGCCCCCTAAGCTGCG 1085
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 GCGGGACCTCCAGGGGAGGCCGTCTTCTTTCTTCGGCATGGACAGGCCCCCTAAGCTGCG 1020

Qy       1086 GGACGAGTTCTGATT 1100
              |||||||||||||||
Db       1021 GGACGAGTTCTGATT 1035
```

RESULT 8
AAD07629
ID AAD07629 standard; cDNA; 915 BP.
XX
AC AAD07629;
XX
DT 10-AUG-2001 (first entry)
XX
DE Human secreted protein-encoding gene 18 cDNA clone HMWEY26, SEQ ID NO:69.
XX
KW Human; secreted protein; proliferative disorder; cancer; tumour;
KW foetal abnormality; developmental abnormality; haematopoietic disorder;
KW immune system disorder; AIDS; autoimmune disease; rheumatoid arthritis;
KW inflammation; allergy; neurological disorder; Alzheimer's disease;
KW Parkinson's disease; cognitive disorder; schizophrenia; asthma;
KW skin disorder; psoriasis; sepsis; diabetes; atherosclerosis;
KW cardiovascular disorder; angiogenic disorder; kidney disorder;
KW gastrointestinal disorder; pregnancy-related disorder; gene therapy;
KW endocrine disorder; infection; wound healing; vulnerary; cell culture;
KW chemotaxis; food additive; binding partner identification; chromosome 3;
KW ss.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT CDS 1. .498
FT /*tag= a
FT /product= "Human secreted protein precursor"
FT /transl_except= (pos:4. .6, aa:Xaa)
FT /transl_except= (pos:10. .12, aa:Xaa)
FT /transl_except= (pos:13. .15, aa:Xaa)
FT /transl_except= (pos:442. .444, aa:Xaa)
FT /note= "Xaa equals any of the naturally occurring L-amino
FT acids; CDS does not include start codon"
FT /partial
FT sig_peptide 1. .3
FT /*tag= b
FT mat_peptide 4. .495
FT /*tag= c
FT /product= "Mature human secreted protein"
XX
PN WO200132676-A1.
XX
PD 10-MAY-2001.
XX
PF 25-OCT-2000; 2000WO-US029365.
XX
PR 29-OCT-1999; 99US-0162237P.
PR 21-JUL-2000; 2000US-0219666P.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Ruben SM, Komatsoulis GA, Shi Y, Olsen HS, Soppet DR;
XX
DR WPI; 2001-328773/34.
DR P-PSDB; AAE03110.
XX
PT Nucleic acids encoding 25 human secreted polypeptides, useful for
PT preventing, diagnosing and/or treating e.g. Gaucher's disease,
PT Alzheimer's disease, Scimitar syndrome, Creutzfeldt-Jacob disease,
PT diabetes mellitus and multiple sclerosis.
XX
PS Claim 1; Page 445; 546pp; English.
XX
CC AAD07571-AAD07645 represent cDNAs corresponding to 25 human secreted
CC protein genes, and AAE03052-AAE03126 represent the proteins they encode.
CC AAE03127-AAE03150 represent human secreted protein fragments. The genes
CC and their corresponding secreted proteins are useful for preventing,
CC treating or ameliorating medical conditions, e.g., by protein or gene
CC therapy. Pathological conditions can be diagnosed by determining the
CC amount of the new protein in a sample or by determining the presence of
CC mutations in the new genes. Specific uses are described for each of the
CC 25 genes, based on the tissues in which they are most highly expressed,
CC and include developing products for the diagnosis or treatment of
CC proliferative disorders, cancer, tumours, foetal and developmental
CC abnormalities, haematopoietic disorders, diseases of the immune system,
CC AIDS, autoimmune diseases (e.g., rheumatoid arthritis), inflammation,
CC allergies, neurological disorders (e.g., Alzheimer's disease, Parkinson's
CC disease), cognitive disorders, schizophrenia, asthma, skin disorders
CC (e.g., psoriasis), sepsis, diabetes, atherosclerosis, cardiovascular
CC disorders, angiogenic disorders, kidney disorders, gastrointestinal
CC disorders, pregnancy-related disorders, endocrine disorders, and
CC infections. The proteins can also be used to aid wound healing and
CC epithelial cell proliferation, to prevent skin aging due to sunburn, to
CC maintain organs before transplantation, for supporting cell culture of
CC primary tissues, to regenerate tissues, to identify their cognate ligands
CC or binding partners, and in chemotaxis, and can be used as a food
CC additive or preservative to modify storage properties. Antibodies
CC specific for a protein of the invention can be used in alleviating
CC symptoms associated with the disorders mentioned above, and in diagnostic
CC immunoassays e.g., radioimmunoassay or enzyme linked immunosorbent assay
CC (ELISA). The present sequence represents a human secreted protein-
CC encoding cDNA of the invention
XX
SQ Sequence 915 BP; 209 A; 253 C; 247 G; 202 T; 0 U; 4 Other;

Query Match 43.0%; Score 473.4; DB 4; Length 915;
Best Local Similarity 98.4%; Pred. No. 8.7e-126;
Matches 477; Conservative 1; Mismatches 7; Indels 0; Gaps 0;

```
Qy        616 CACCTTGTCCGGCCCTATGTGGGCATTCATCTGCGCATTGGCTCTGACTGGAAGAACGCC 675
```

```
Db         16 CACTTTGCCCGAACCTACCCGGGCATTCATCTGCGCATTGGCTCTGACTGGAAGAACGCC 75

Qy        676 TGTGCCATGCTGAAGGACGGGACTGCAGGCTCGCACTTCATGGCCTCTCCGCAGTGTGTG 735
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         76 TGTGCCATGCTGAAGGACGGGACTGCAGGCTCGCACTTCATGGCCTCTCCGCAGTGTGTG 135

Qy        736 GGCTACAGCCGCAGCACAGCGGCCCCCCTCACGATGACTATGTGCCTGCCTGACCTGAAG 795
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        136 GGCTACAGCCGCAGCACAGCGGCCCCCCTCACGATGACTATGTGCCTGCCTGACCTGAAG 195

Qy        796 GAGATCCAGAGGGCTGTGAAGCTCTGGGTGAGGTCGCTGGATGCCCAGTCGGTCTACGTT 855
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        196 GAGATCCAGAGGGCTGTGAAGCTCTGGGTGAGGTCGCTGGATGCCCAGTCGGTCTACGTT 255

Qy        856 GCTACTGATTCCGAGAGTTATGTGCCTGAGCTCCAACAGCTCTTCAAAGGGAAGGTGAAG 915
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        256 GCTACTGATTCCGAGAGTTATGTGCCTGAGCTCCAACAGCTCTTCAAAGGGAAGGTGAAG 315

Qy        916 GTGGTGAGCCTGAAGCCTGAGGTGGCCCAGGTCGACCTGTACATCCTCGGCCAAGCCGAC 975
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        316 GTGGTGAGCCTGAAGCCTGAGGTGGCCCAGGTCGACCTGTACATCCTCGGCCAAGCCGAC 375

Qy        976 CACTTTATTGGCAACTGTGTCTCCTCCTTCACTGCCTTTGTGAAGCGGGAGCGGGACCTC 1035
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        376 CACTTTATTGGCAACTGTGTCTCCTCCTTCACTGCCTTTGTGAAGCGGGAGCGGGACCTC 435

Qy       1036 CAGGGGAGGCCGTCTTCTTTCTTCGGCATGGACAGGCCCCCTAAGCTGCGGGACGAGTTC 1095
              ||||||||:|||||||||||||||||||||||||||||||||||||||||||||||||||
Db        436 CAGGGGAGSCCGTCTTCTTTCTTCGGCATGGACAGGCCCCCTAAGCTGCGGGACGAGTTC 495

Qy       1096 TGATT 1100
              |||||
Db        496 TGATT 500
```

RESULT 9
ADQ56824
ID ADQ56824 standard; DNA; 610 BP.
XX
AC ADQ56824;
XX
DT 21-OCT-2004 (first entry)
XX
DE Novel canine microarray-related DNA sequence SeqID8126.
XX
KW canine microarray; drug screening; toxicity assay;
KW environmental pollutant; cellular response; gene expression profile;
KW toxic response; liver necrosis; fatty liver disease;
KW protein adduct formation; hepatitis; dog; ds.
XX
OS Canis familiaris.
XX
PN WO2004063324-A2.
XX
PD 29-JUL-2004.
XX
PF 05-MAY-2003; 2003WO-US013853.
XX
PR 03-MAY-2002; 2002US-0377240P.
XX
PA (GENE-) GENE LOGIC INC.
PA (PFIZ) PFIZER PROD INC.
XX
PI Diggans JC, Porter M, Wei T;
XX
DR WPI; 2004-561890/54.
XX
PT New isolated nucleic acid molecule, useful for drug screening and
PT toxicity assays or for assessing the impact, including toxicity, of a
PT compound, pharmaceutical agent or environmental pollutant on a cell or
PT living organism.
XX
PS Claim 1; SEQ ID NO 8126; 41pp; English..
XX
CC This invention is related to a novel isolated canine nucleic acid
CC sequences and the construction of canine microarrays containing a
CC significant portion of the canine genome. The isolated canine nucleic
CC acid sequences of the invention may be useful for drug screening and
CC toxicity assays. The invention is therefore useful for assessing the
CC impact, including toxicity, of a compound, pharmaceutical agent or
CC environmental pollutant on a cell or living organism. The methods are
CC useful for detecting genes that are up- or down-regulated in canines in a
CC disease state. The sequences are useful as diagnostic agents or markers
CC to detect a cellular response in a sample individually or as part of a
CC gene expression profile. It is also useful as a target for agents that
CC modulate gene expression or activity. The database is useful for
CC producing electronic Northerns that allow the user to determine the cell
CC type or tissue in which a given gene is expressed and to allow
CC determination of the abundance or expression level of a given gene in a
CC particular tissue or cell. The methods are useful for determining the
CC similarity of a toxic response to one or more individual compounds. The
CC methods are useful for predicting at least one toxic response or the
CC likelihood that a compound or test agent will induce various specific
CC pathologies such as those of the liver (liver necrosis, fatty liver
CC disease, protein adduct formation or hepatitis), those of the kidney,
CC heart, brain or testes, or other pathologies associated with at least one
CC of the toxins. The methods are also useful for predicting or elucidating
CC the potential cellular pathways influenced, induced or modulated by the
CC compound or test agent due to the similarity of the expression profile
CC compared to the profile induced by a known toxin. The present sequence is
CC that of a canine DNA sequence which was claimed for use during the
CC production of a canine microarray of the invention.
XX
SQ Sequence 610 BP; 155 A; 171 C; 149 G; 129 T; 0 U; 6 Other;

Query Match 36.9%; Score 406.2; DB 13; Length 610;
Best Local Similarity 89.0%; Pred. No. 1.8e-106;
Matches 446; Conservative 0; Mismatches 54; Indels 1; Gaps 1;

```
Qy        171 CACCAACCTCCATGTGTCCTACCAGAAGTACTTCAAGCTGGAGCCCCCTCCAGGCTTACCA 230
              |||||| |||||||||| |||||||||||||||||||||  |||||||| |||  |||||
Db          1 CACCAATCTCCATGTGTNCTACCAGAAGTACTTCAAGCTNNAGCCCCTCNAGGNNTACCA 60

Qy        231 TCGGGTCATCAGCTTGGAGGATTTCATGGAGAAGCTGGCACCCACCCACTGGCCCCCTGA 290
              |||||| |||||| | ||||| |||||||||||||||||||||| |||||||||| ||
Db         61 TCGGGTTATCAGCCTAGAGGACTTCATGGAGAAGCTGGCACCCACTCACTGGCCCCCCGA 120

Qy        291 GAAGCGGGTGGCATACTGCTTTGAGGTGGCAGCCCAGCGAAGCCCAGATAAGAAGACGTG 350
              || ||| |||||||||||||||| ||||||||||||||||||||| |||||||||| ||
Db        121 GAGGCGAGTGGCATACTGCTTTGAAGTGGCAGCCCAGCGAAGCCCTGATAAGAAGACATG 180

Qy        351 CCCCATGAAGGAAGGAAACCCCTTTGGCCCATTCTGGGATCAGTTTCATGTGAGTTTCAA 410
              |||||||||||||||||| |||||||||||||| |||||||||||||||||||||||||||
Db        181 CCCCATGAAGGAAGGAAATCCCTTTGGCCCATTTTGGGATCAGTTTCATGTGAGTTTCAA 240

Qy        411 CAAGTCGGAGCTTTTTACAGGCATTTCCTTCAGTGCTTCCTACAGAGAACAATGGAGCCA 470
              ||||||  |||||||| |||||||||||||||||| ||||||| ||| |||||||  ||
Db        241 CAAGTCCAAGCTTTTTTCAGGCATTTCCTTCAGTGCCTCCTACAAAGACCAATGGATTCA 300

Qy        471 GAGATTTTCTCCAAAGGAACATCCGGTGCTTGCCCTGCCAGGAGCCCCAGCCCAGTTCCC 530
              |||||||||||||||||||||||| ||||| ||||  || |||||||| ||||||||||
Db        301 GAGATTTTCTCCAAAGGAACATCCAGTGCTGGCCC-CCCGGGAGCCCCCGCCCAGTTCCC 359

Qy        531 CGTCCTAGAAGAACACAGGCCACTACAGAAGTACATGGTATGGTCAGACGAAATGGTGAA 590
              |||||| || || ||||| ||||| |||||||| |||||||||||| || |||||||
Db        360 CGTCCTGGAGGAGCACAGACCACTCCAGAAGTATGTGGTATGGTCAGATGAGATGGTGAG 419

Qy        591 GACGGGAGAGGCCCAGATTCATGCCCACCTTGTCCGGCCCTATGTGGGCATTCATCTGCG 650
              |||||||||||||||| ||||   |||||||  ||||||| || |||||||||| |||||
Db        420 GACGGGAGAGGCCCACATTCTCACCCACCTCATCCGGCCGTACGTGGGCATTCACCTGCG 479

Qy        651 CATTGGCTCTGACTGGAAGAA 671
              |||||||||||||||| | |
Db        480 CATTGGCTCTGACTGGATGGA 500
```

```
RESULT 10
AAK79915
ID   AAK79915 standard; DNA; 10331 BP.
XX
AC   AAK79915;
XX
DT   07-NOV-2001  (first entry)
XX
DE   Human immune/haematopoietic antigen genomic sequence SEQ ID NO:34727.
XX
KW   Human; immune; haematopoietic; immune/haematopoietic antigen; cancer;
KW   cytostatic; gene therapy; vaccine; metastasis; ds.
XX
OS   Homo sapiens.
XX
PN   WO200157182-A2.
XX
PD   09-AUG-2001.
XX
PF   17-JAN-2001; 2001WO-US001354.
XX
PR   31-JAN-2000; 2000US-0179065P.
PR   04-FEB-2000; 2000US-0180628P.
PR   24-FEB-2000; 2000US-0184664P.
PR   02-MAR-2000; 2000US-0186350P.
PR   16-MAR-2000; 2000US-0189874P.
PR   17-MAR-2000; 2000US-0190076P.
PR   18-APR-2000; 2000US-0198123P.
PR   19-MAY-2000; 2000US-0205515P.
PR   07-JUN-2000; 2000US-0209467P.
PR   28-JUN-2000; 2000US-0214886P.
PR   30-JUN-2000; 2000US-0215135P.
PR   07-JUL-2000; 2000US-0216647P.
PR   07-JUL-2000; 2000US-0216880P.
PR   11-JUL-2000; 2000US-0217487P.
PR   11-JUL-2000; 2000US-0217496P.
PR   14-JUL-2000; 2000US-0218290P.
PR   26-JUL-2000; 2000US-0220963P.
PR   26-JUL-2000; 2000US-0220964P.
PR   14-AUG-2000; 2000US-0224518P.
PR   14-AUG-2000; 2000US-0224519P.
PR   14-AUG-2000; 2000US-0225213P.
PR   14-AUG-2000; 2000US-0225214P.
PR   14-AUG-2000; 2000US-0225266P.
PR   14-AUG-2000; 2000US-0225267P.
PR   14-AUG-2000; 2000US-0225268P.
PR   14-AUG-2000; 2000US-0225270P.
PR   14-AUG-2000; 2000US-0225447P.
PR   14-AUG-2000; 2000US-0225757P.
PR   14-AUG-2000; 2000US-0225758P.
PR   14-AUG-2000; 2000US-0225759P.
PR   18-AUG-2000; 2000US-0226279P.
PR   22-AUG-2000; 2000US-0226681P.
PR   22-AUG-2000; 2000US-0226868P.
PR   22-AUG-2000; 2000US-0227182P.
PR   23-AUG-2000; 2000US-0227009P.
PR   30-AUG-2000; 2000US-0228924P.
PR   01-SEP-2000; 2000US-0229287P.
PR   01-SEP-2000; 2000US-0229343P.
PR   01-SEP-2000; 2000US-0229344P.
PR   01-SEP-2000; 2000US-0229345P.
PR   05-SEP-2000; 2000US-0229509P.
PR   05-SEP-2000; 2000US-0229513P.
PR   06-SEP-2000; 2000US-0230437P.
PR   06-SEP-2000; 2000US-0230438P.
PR   08-SEP-2000; 2000US-0231242P.
PR   08-SEP-2000; 2000US-0231243P.
PR   08-SEP-2000; 2000US-0231244P.
PR   08-SEP-2000; 2000US-0231413P.
PR   08-SEP-2000; 2000US-0231414P.
PR   08-SEP-2000; 2000US-0232080P.
PR   08-SEP-2000; 2000US-0232081P.
PR   12-SEP-2000; 2000US-0231968P.
PR   14-SEP-2000; 2000US-0232397P.
PR   14-SEP-2000; 2000US-0232398P.
PR   14-SEP-2000; 2000US-0232399P.
PR   14-SEP-2000; 2000US-0232400P.
PR   14-SEP-2000; 2000US-0232401P.
PR   14-SEP-2000; 2000US-0233063P.
PR   14-SEP-2000; 2000US-0233064P.
PR   14-SEP-2000; 2000US-0233065P.
PR   21-SEP-2000; 2000US-0234223P.
PR   21-SEP-2000; 2000US-0234274P.
PR   25-SEP-2000; 2000US-0234997P.
PR   25-SEP-2000; 2000US-0234998P.
PR   26-SEP-2000; 2000US-0235484P.
PR   27-SEP-2000; 2000US-0235834P.
PR   27-SEP-2000; 2000US-0235836P.
PR   29-SEP-2000; 2000US-0236327P.
PR   29-SEP-2000; 2000US-0236367P.
PR   29-SEP-2000; 2000US-0236368P.
PR   29-SEP-2000; 2000US-0236369P.
PR   29-SEP-2000; 2000US-0236370P.
PR   02-OCT-2000; 2000US-0236802P.
PR   02-OCT-2000; 2000US-0237037P.
PR   02-OCT-2000; 2000US-0237038P.
PR   02-OCT-2000; 2000US-0237039P.
PR   02-OCT-2000; 2000US-0237040P.
PR   13-OCT-2000; 2000US-0239935P.
PR   13-OCT-2000; 2000US-0239937P.
PR   20-OCT-2000; 2000US-0240960P.
PR   20-OCT-2000; 2000US-0241221P.
PR   20-OCT-2000; 2000US-0241785P.
PR   20-OCT-2000; 2000US-0241786P.
PR   20-OCT-2000; 2000US-0241787P.
PR   20-OCT-2000; 2000US-0241808P.
PR   20-OCT-2000; 2000US-0241809P.
PR   20-OCT-2000; 2000US-0241826P.
PR   01-NOV-2000; 2000US-0244617P.
PR   08-NOV-2000; 2000US-0246474P.
PR   08-NOV-2000; 2000US-0246475P.
PR   08-NOV-2000; 2000US-0246476P.
PR   08-NOV-2000; 2000US-0246477P.
PR   08-NOV-2000; 2000US-0246478P.
PR   08-NOV-2000; 2000US-0246523P.
PR   08-NOV-2000; 2000US-0246524P.
PR   08-NOV-2000; 2000US-0246525P.
PR   08-NOV-2000; 2000US-0246526P.
PR   08-NOV-2000; 2000US-0246527P.
PR   08-NOV-2000; 2000US-0246528P.
PR   08-NOV-2000; 2000US-0246532P.
PR   08-NOV-2000; 2000US-0246609P.
PR   08-NOV-2000; 2000US-0246610P.
PR   08-NOV-2000; 2000US-0246611P.
PR   08-NOV-2000; 2000US-0246613P.
PR   17-NOV-2000; 2000US-0249207P.
PR   17-NOV-2000; 2000US-0249208P.
PR   17-NOV-2000; 2000US-0249209P.
PR   17-NOV-2000; 2000US-0249210P.
PR   17-NOV-2000; 2000US-0249211P.
PR   17-NOV-2000; 2000US-0249212P.
PR   17-NOV-2000; 2000US-0249213P.
PR   17-NOV-2000; 2000US-0249214P.
PR   17-NOV-2000; 2000US-0249215P.
PR   17-NOV-2000; 2000US-0249216P.
PR   17-NOV-2000; 2000US-0249217P.
PR   17-NOV-2000; 2000US-0249218P.
PR   17-NOV-2000; 2000US-0249244P.
PR   17-NOV-2000; 2000US-0249245P.
PR   17-NOV-2000; 2000US-0249264P.
PR   17-NOV-2000; 2000US-0249265P.
PR   17-NOV-2000; 2000US-0249297P.
PR   17-NOV-2000; 2000US-0249299P.
PR   17-NOV-2000; 2000US-0249300P.
PR   01-DEC-2000; 2000US-0250160P.
```

```
PR   01-DEC-2000; 2000US-0250391P.
PR   05-DEC-2000; 2000US-0251030P.
PR   05-DEC-2000; 2000US-0251988P.
PR   05-DEC-2000; 2000US-0256719P.
PR   06-DEC-2000; 2000US-0251479P.
PR   08-DEC-2000; 2000US-0251856P.
PR   08-DEC-2000; 2000US-0251868P.
PR   08-DEC-2000; 2000US-0251869P.
PR   08-DEC-2000; 2000US-0251989P.
PR   08-DEC-2000; 2000US-0251990P.
PR   11-DEC-2000; 2000US-0254097P.
PR   05-JAN-2001; 2001US-0259678P.
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
XX
PI   Rosen CA,  Barash SC,  Ruben SM;
XX
DR   WPI; 2001-483426/52.
XX
PT   Nucleic acids encoding human immune/hematopoietic antigen polypeptides,
PT   useful for preventing, diagnosing and/or treating cancers and metastasis.
XX
PS   Disclosure; SEQ ID NO 34727; 3071pp + Sequence Listing; English.
XX
CC   AAK54951 to AAK64702 encode the human immune/haematopoietic antigen (I)
CC   amino acid sequences given in AAM82170 to AAM91921. (I) have cytostatic
CC   activity, and can be used in gene therapy and vaccine production. (I)
CC   proteins and polynucleotides may be used in the prevention, diagnosis and
CC   treatment of diseases associated with inappropriate (I) expression. For
CC   example, they may be used to treat disorders associated with decreased
CC   expression by rectifying mutations or deletions in a patient's genome
CC   that affect the activity of (I) by expressing inactive proteins or to
CC   supplement the patients own production of (I). Additionally, (I)
CC   polynucleotides may be used to produce the secreted (I), by inserting the
CC   nucleic acids into a host cell and culturing the cell to express the
CC   protein. (I) proteins and polynucleotides may be used to prevent,
CC   diagnose and treat immune/haematopoietic-related diseases, especially
CC   cancers and cancer metastases of haematopoietic-derived cells. AAK64703
CC   to AAK87694 represent human immune/haematopoietic antigen genomic
CC   sequences from the present invention. AAK54942 to AAK54950 and AAM82169
CC   represent sequences used in the exemplification of the present invention
XX
SQ   Sequence 10331 BP; 2718 A; 2405 C; 2599 G; 2609 T; 0 U; 0 Other;

  Query Match             22.8%;  Score 251.2;  DB 4;  Length 10331;
  Best Local Similarity   97.0%;  Pred. No. 3.4e-61;
  Matches  256;  Conservative    0;  Mismatches    8;  Indels    0;  Gaps    0;

Qy        656 GCTCTGACTGGAAGAACGCCTGTGCCATGCTGAAGGACGGGACTGCAGGCTCGCACTTCA 715
              |||||   | ||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2480 GCTCTCCGTAGAAGAACGCCTGTGCCATGCTGAAGGACGGGACTGCAGGCTCGCACTTCA 2539

Qy        716 TGGCCTCTCCGCAGTGTGTGGGCTACAGCCGCAGCACAGCGGCCCCCCTCACGATGACTA 775
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2540 TGGCCTCTCCGCAGTGTGTGGGCTACAGCCGCAGCACAGCGGCCCCCCTCACGATGACTA 2599

Qy        776 TGTGCCTGCCTGACCTGAAGGAGATCCAGAGGGCTGTGAAGCTCTGGGTGAGGTCGCTGG 835
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2600 TGTGCCTGCCTGACCTGAAGGAGATCCAGAGGGCTGTGAAGCTCTGGGTGAGGTCGCTGG 2659

Qy        836 ATGCCCAGTCGGTCTACGTTGCTACTGATTCCGAGAGTTATGTGCCTGAGCTCCAACAGC 895
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2660 ATGCCCAGTCGGTCTACGTTGCTACTGATTCCGAGAGTTATGTGCCTGAGCTCCAACAGC 2719

Qy        896 TCTTCAAAGGGAAGGTGAAGGTGG 919
              |||||||||||||||    ||||
Db       2720 TCTTCAAAGGGAAGGTATGTGTGG 2743


RESULT 11
AAL03344
ID   AAL03344 standard; DNA; 10331 BP.
XX
AC   AAL03344;
XX
DT   21-NOV-2001  (first entry)
XX
DE   Human reproductive system related antigen DNA SEQ ID NO: 6032.
XX
KW   Human; reproductive system related antigen; reproductive system disorder;
KW   cancer; gene therapy; ds.
XX
OS   Homo sapiens.
XX
PN   WO200155320-A2.
XX
PD   02-AUG-2001.
XX
PF   17-JAN-2001; 2001WO-US001339.
XX
PR   31-JAN-2000; 2000US-0179065P.
PR   04-FEB-2000; 2000US-0180628P.
PR   24-FEB-2000; 2000US-0184664P.
PR   02-MAR-2000; 2000US-0186350P.
PR   16-MAR-2000; 2000US-0189874P.
PR   17-MAR-2000; 2000US-0190076P.
PR   18-APR-2000; 2000US-0198123P.
PR   19-MAY-2000; 2000US-0205515P.
PR   07-JUN-2000; 2000US-0209467P.
PR   28-JUN-2000; 2000US-0214886P.
PR   30-JUN-2000; 2000US-0215135P.
PR   07-JUL-2000; 2000US-0216647P.
PR   07-JUL-2000; 2000US-0216880P.
PR   11-JUL-2000; 2000US-0217487P.
PR   11-JUL-2000; 2000US-0217496P.
PR   14-JUL-2000; 2000US-0218290P.
PR   26-JUL-2000; 2000US-0220963P.
PR   26-JUL-2000; 2000US-0220964P.
PR   14-AUG-2000; 2000US-0224518P.
PR   14-AUG-2000; 2000US-0224519P.
PR   14-AUG-2000; 2000US-0225213P.
PR   14-AUG-2000; 2000US-0225214P.
PR   14-AUG-2000; 2000US-0225266P.
PR   14-AUG-2000; 2000US-0225267P.
PR   14-AUG-2000; 2000US-0225268P.
PR   14-AUG-2000; 2000US-0225270P.
PR   14-AUG-2000; 2000US-0225447P.
PR   14-AUG-2000; 2000US-0225757P.
PR   14-AUG-2000; 2000US-0225758P.
PR   14-AUG-2000; 2000US-0225759P.
PR   18-AUG-2000; 2000US-0226279P.
PR   22-AUG-2000; 2000US-0226681P.
PR   22-AUG-2000; 2000US-0226868P.
PR   22-AUG-2000; 2000US-0227182P.
PR   23-AUG-2000; 2000US-0227009P.
PR   30-AUG-2000; 2000US-0228924P.
PR   01-SEP-2000; 2000US-0229287P.
PR   01-SEP-2000; 2000US-0229343P.
PR   01-SEP-2000; 2000US-0229344P.
PR   01-SEP-2000; 2000US-0229345P.
PR   05-SEP-2000; 2000US-0229509P.
PR   05-SEP-2000; 2000US-0229513P.
PR   06-SEP-2000; 2000US-0230437P.
PR   06-SEP-2000; 2000US-0230438P.
PR   08-SEP-2000; 2000US-0231242P.
PR   08-SEP-2000; 2000US-0231243P.
PR   08-SEP-2000; 2000US-0231244P.
PR   08-SEP-2000; 2000US-0231413P.
PR   08-SEP-2000; 2000US-0231414P.
PR   08-SEP-2000; 2000US-0232080P.
PR   08-SEP-2000; 2000US-0232081P.
PR   12-SEP-2000; 2000US-0231968P.
PR   14-SEP-2000; 2000US-0232397P.
PR   14-SEP-2000; 2000US-0232398P.
PR   14-SEP-2000; 2000US-0232399P.
PR   14-SEP-2000; 2000US-0232400P.
```

PR 14-SEP-2000; 2000US-0232401P.
PR 14-SEP-2000; 2000US-0233063P.
PR 14-SEP-2000; 2000US-0233064P.
PR 14-SEP-2000; 2000US-0233065P.
PR 21-SEP-2000; 2000US-0234223P.
PR 21-SEP-2000; 2000US-0234274P.
PR 25-SEP-2000; 2000US-0234997P.
PR 25-SEP-2000; 2000US-0234998P.
PR 26-SEP-2000; 2000US-0235484P.
PR 27-SEP-2000; 2000US-0235834P.
PR 27-SEP-2000; 2000US-0235836P.
PR 29-SEP-2000; 2000US-0236327P.
PR 29-SEP-2000; 2000US-0236367P.
PR 29-SEP-2000; 2000US-0236368P.
PR 29-SEP-2000; 2000US-0236369P.
PR 29-SEP-2000; 2000US-0236370P.
PR 02-OCT-2000; 2000US-0236802P.
PR 02-OCT-2000; 2000US-0237037P.
PR 02-OCT-2000; 2000US-0237038P.
PR 02-OCT-2000; 2000US-0237039P.
PR 02-OCT-2000; 2000US-0237040P.
PR 13-OCT-2000; 2000US-0239935P.
PR 13-OCT-2000; 2000US-0239937P.
PR 20-OCT-2000; 2000US-0240960P.
PR 20-OCT-2000; 2000US-0241221P.
PR 20-OCT-2000; 2000US-0241785P.
PR 20-OCT-2000; 2000US-0241786P.
PR 20-OCT-2000; 2000US-0241787P.
PR 20-OCT-2000; 2000US-0241808P.
PR 20-OCT-2000; 2000US-0241809P.
PR 20-OCT-2000; 2000US-0241826P.
PR 01-NOV-2000; 2000US-0244617P.
PR 08-NOV-2000; 2000US-0246474P.
PR 08-NOV-2000; 2000US-0246475P.
PR 08-NOV-2000; 2000US-0246476P.
PR 08-NOV-2000; 2000US-0246477P.
PR 08-NOV-2000; 2000US-0246478P.
PR 08-NOV-2000; 2000US-0246523P.
PR 08-NOV-2000; 2000US-0246524P.
PR 08-NOV-2000; 2000US-0246525P.
PR 08-NOV-2000; 2000US-0246526P.
PR 08-NOV-2000; 2000US-0246527P.
PR 08-NOV-2000; 2000US-0246528P.
PR 08-NOV-2000; 2000US-0246532P.
PR 08-NOV-2000; 2000US-0246609P.
PR 08-NOV-2000; 2000US-0246610P.
PR 08-NOV-2000; 2000US-0246611P.
PR 08-NOV-2000; 2000US-0246613P.
PR 17-NOV-2000; 2000US-0249207P.
PR 17-NOV-2000; 2000US-0249208P.
PR 17-NOV-2000; 2000US-0249209P.
PR 17-NOV-2000; 2000US-0249210P.
PR 17-NOV-2000; 2000US-0249211P.
PR 17-NOV-2000; 2000US-0249212P.
PR 17-NOV-2000; 2000US-0249213P.
PR 17-NOV-2000; 2000US-0249214P.
PR 17-NOV-2000; 2000US-0249215P.
PR 17-NOV-2000; 2000US-0249216P.
PR 17-NOV-2000; 2000US-0249217P.
PR 17-NOV-2000; 2000US-0249218P.
PR 17-NOV-2000; 2000US-0249244P.
PR 17-NOV-2000; 2000US-0249245P.
PR 17-NOV-2000; 2000US-0249264P.
PR 17-NOV-2000; 2000US-0249265P.
PR 17-NOV-2000; 2000US-0249297P.
PR 17-NOV-2000; 2000US-0249299P.
PR 17-NOV-2000; 2000US-0249300P.
PR 01-DEC-2000; 2000US-0250160P.
PR 01-DEC-2000; 2000US-0250391P.
PR 05-DEC-2000; 2000US-0251030P.
PR 05-DEC-2000; 2000US-0251988P.
PR 05-DEC-2000; 2000US-0256719P.
PR 06-DEC-2000; 2000US-0251479P.
PR 08-DEC-2000; 2000US-0251856P.
PR 08-DEC-2000; 2000US-0251868P.
PR 08-DEC-2000; 2000US-0251869P.
PR 08-DEC-2000; 2000US-0251989P.
PR 08-DEC-2000; 2000US-0251990P.
PR 11-DEC-2000; 2000US-0254097P.
PR 05-JAN-2001; 2001US-0259678P.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Rosen CA, Barash SC, Ruben SM;
XX
DR WPI; 2001-465570/50.
XX
PT Isolated nucleic acid molecule encoding a reproductive system antigen is
PT used in preventing, treating or ameliorating a medical condition.
XX
PS Disclosure; SEQ ID NO 6032; 1297pp + Sequence Listing; English.
XX
CC The present invention provides the protein and coding sequences of a
CC number of human reproductive system related antigens. These can be used
CC in the prevention and treatment of reproductive system disorders,
CC including cancer. The present sequence is a genomic sequence encoding a
CC protein of the invention
XX
SQ Sequence 10331 BP; 2718 A; 2405 C; 2599 G; 2609 T; 0 U; 0 Other;

Query Match 22.8%; Score 251.2; DB 4; Length 10331;
Best Local Similarity 97.0%; Pred. No. 3.4e-61;
Matches 256; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

```
Qy       656 GCTCTGACTGGAAGAACGCCTGTGCCATGCTGAAGGACGGGACTGCAGGCTCGCACTTCA 715
            |||||   | ||||||||||||||||||||||||||||||||||||||||||||||||||
Db      2480 GCTCTCCGTAGAAGAACGCCTGTGCCATGCTGAAGGACGGGACTGCAGGCTCGCACTTCA 2539

Qy       716 TGGCCTCTCCGCAGTGTGTGGGCTACAGCCGCAGCACAGCGGCCCCCCTCACGATGACTA 775
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      2540 TGGCCTCTCCGCAGTGTGTGGGCTACAGCCGCAGCACAGCGGCCCCCCTCACGATGACTA 2599

Qy       776 TGTGCCTGCCTGACCTGAAGGAGATCCAGAGGGCTGTGAAGCTCTGGGTGAGGTCGCTGG 835
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      2600 TGTGCCTGCCTGACCTGAAGGAGATCCAGAGGGCTGTGAAGCTCTGGGTGAGGTCGCTGG 2659

Qy       836 ATGCCCAGTCGGTCTACGTTGCTACTGATTCCGAGAGTTATGTGCCTGAGCTCCAACAGC 895
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      2660 ATGCCCAGTCGGTCTACGTTGCTACTGATTCCGAGAGTTATGTGCCTGAGCTCCAACAGC 2719

Qy       896 TCTTCAAAGGGAAGGTGAAGGTGG 919
            |||||||||||||||    ||||
Db      2720 TCTTCAAAGGGAAGGTATGTGTGG 2743
```

RESULT 12
ACH32225
ID ACH32225 standard; cDNA; 479 BP.
XX
AC ACH32225;
XX
DT 13-OCT-2003 (first entry)
XX
DE Human endothelial cell cDNA #358.
XX
KW Human; ss; sequencing by hybridisation; SBH; expressed sequence tag; EST;
KW genome mapping; biodiversity; genetic disorder.
XX
OS Homo sapiens.
XX
PN US2003073623-A1.
XX
PD 17-APR-2003.
XX
PF 30-JUL-2001; 2001US-00918995.
XX

PR 30-JUL-2001; 2001US-00918995.
XX
PA (DRMA/) DRMANAC R T.
PA (LABA/) LABAT I.
PA (STAC/) STACHE-CRAIN B.
PA (DICK/) DICKSON M C.
PA (JONE/) JONES L W.
XX
PI Drmanac RT, Labat I, Stache-Crain B, Dickson MC, Jones LW;
XX
DR WPI; 2003-615964/58.
XX
PT New polynucleotide sequences obtained from various cDNA libraries, useful
PT as hybridization probes, as oligomers for PCR, for chromosome and gene
PT mapping, in the recombinant production of protein, or in generating
PT antisense DNA or RNA.
XX
PS Claim 1; SEQ ID NO 19437; 44pp; English.
XX
CC The invention relates to an isolated polynucleotide comprising any one of
CC 38043 cDNA sequences, appearing as ACH12789-ACH50831, whose sequence was
CC determined by the technique of SBH (sequencing by hybridisation). Also
CC included is a purified polypeptide comprising a sequence corresponding to
CC a reading frame of the novel polynucleotide. The nucleic acid sequences
CC are useful in diagnostics as expressed sequence tags (EST) for
CC identifying expressed genes or for physical mapping of the human genome,
CC in forensics, in assessing biodiversities, or in identifying mutations
CC responsible for genetic disorders and other traits. The nucleotide
CC sequences are also useful as hybridisation probes, as oligomers for PCR,
CC for chromosome and gene mapping, in the recombinant production of
CC protein, or in generating antisense DNA or RNA. The purified polypeptide
CC is useful for generating antibodies specific for it. The present sequence
CC is one of the 38043 isolated cDNA/EST sequences. Note: The sequence data
CC for this patent did not form part of the printed specification, but was
CC obtained in electronic format directly from USPTO at
CC seqdata.uspto.gov/sequence.html?DocID=20030073623
XX
SQ Sequence 479 BP; 99 A; 140 C; 143 G; 97 T; 0 U; 0 Other;

Query Match 21.2%; Score 233; DB 9; Length 479;
Best Local Similarity 100.0%; Pred. No. 1.5e-56;
Matches 233; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy        868 GAGAGTTATGTGCCTGAGCTCCAACAGCTCTTCAAAGGGAAGGTGAAGGTGGTGAGCCTG 927
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         31 GAGAGTTATGTGCCTGAGCTCCAACAGCTCTTCAAAGGGAAGGTGAAGGTGGTGAGCCTG 90

Qy        928 AAGCCTGAGGTGGCCCAGGTCGACCTGTACATCCTCGGCCAAGCCGACCACTTTATTGGC 987
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         91 AAGCCTGAGGTGGCCCAGGTCGACCTGTACATCCTCGGCCAAGCCGACCACTTTATTGGC 150

Qy        988 AACTGTGTCTCCTCCTTCACTGCCTTTGTGAAGCGGGAGCGGGACCTCCAGGGGAGGCCG 1047
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        151 AACTGTGTCTCCTCCTTCACTGCCTTTGTGAAGCGGGAGCGGGACCTCCAGGGGAGGCCG 210

Qy       1048 TCTTCTTTCTTCGGCATGGACAGGCCCCCTAAGCTGCGGGACGAGTTCTGATT 1100
              ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        211 TCTTCTTTCTTCGGCATGGACAGGCCCCCTAAGCTGCGGGACGAGTTCTGATT 263
```

RESULT 13
ACH35532
ID ACH35532 standard; cDNA; 477 BP.
XX
AC ACH35532;
XX
DT 13-OCT-2003 (first entry)
XX
DE Human endothelial cell cDNA #3665.
XX
KW Human; ss; sequencing by hybridisation; SBH; expressed sequence tag; EST;
KW genome mapping; biodiversity; genetic disorder.
XX
OS Homo sapiens.
XX
PN US2003073623-A1.
XX
PD 17-APR-2003.
XX
PF 30-JUL-2001; 2001US-00918995.
XX
PR 30-JUL-2001; 2001US-00918995.
XX
PA (DRMA/) DRMANAC R T.
PA (LABA/) LABAT I.
PA (STAC/) STACHE-CRAIN B.
PA (DICK/) DICKSON M C.
PA (JONE/) JONES L W.
XX
PI Drmanac RT, Labat I, Stache-Crain B, Dickson MC, Jones LW;
XX
DR WPI; 2003-615964/58.
XX
PT New polynucleotide sequences obtained from various cDNA libraries, useful
PT as hybridization probes, as oligomers for PCR, for chromosome and gene
PT mapping, in the recombinant production of protein, or in generating
PT antisense DNA or RNA.
XX
PS Claim 1; SEQ ID NO 22744; 44pp; English.
XX
CC The invention relates to an isolated polynucleotide comprising any one of
CC 38043 cDNA sequences, appearing as ACH12789-ACH50831, whose sequence was
CC determined by the technique of SBH (sequencing by hybridisation). Also
CC included is a purified polypeptide comprising a sequence corresponding to
CC a reading frame of the novel polynucleotide. The nucleic acid sequences
CC are useful in diagnostics as expressed sequence tags (EST) for
CC identifying expressed genes or for physical mapping of the human genome,
CC in forensics, in assessing biodiversities, or in identifying mutations
CC responsible for genetic disorders and other traits. The nucleotide
CC sequences are also useful as hybridisation probes, as oligomers for PCR,
CC for chromosome and gene mapping, in the recombinant production of
CC protein, or in generating antisense DNA or RNA. The purified polypeptide
CC is useful for generating antibodies specific for it. The present sequence
CC is one of the 38043 isolated cDNA/EST sequences. Note: The sequence data
CC for this patent did not form part of the printed specification, but was
CC obtained in electronic format directly from USPTO at
CC seqdata.uspto.gov/sequence.html?DocID=20030073623
XX
SQ Sequence 477 BP; 95 A; 141 C; 142 G; 97 T; 0 U; 2 Other;

Query Match 20.2%; Score 221.8; DB 9; Length 477;
Best Local Similarity 96.3%; Pred. No. 2.5e-53;
Matches 237; Conservative 0; Mismatches 8; Indels 1; Gaps 1;

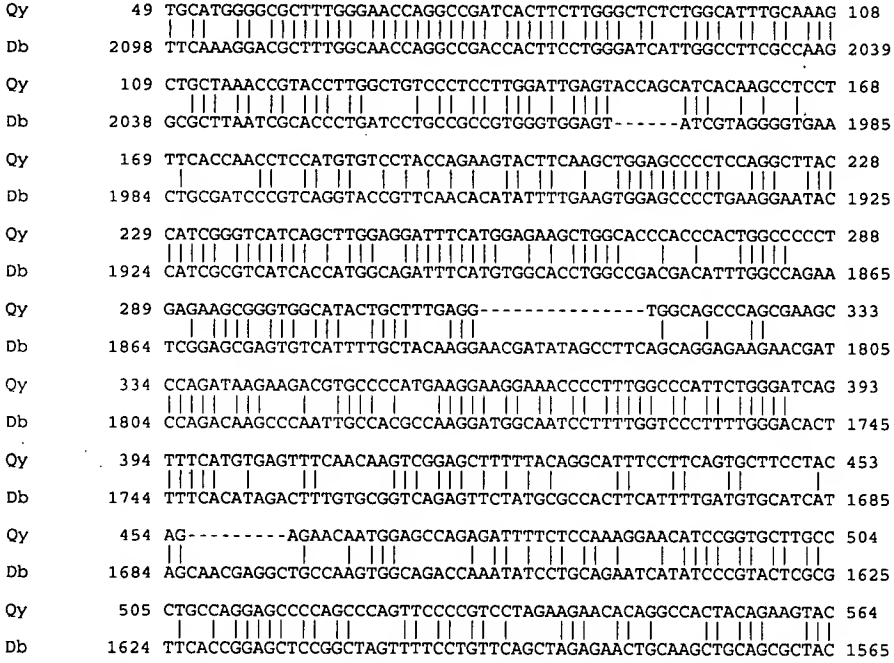
```
Qy        855 TGCTACTGATTCCGAGAGTTATGTGCCTGAGCTCCAACAGCTCTTCAAAGGGAAGGTGAA 914
              || | ||||   | |||| ||||||||||||||| |||| ||||||||||||||||||
Db         11 TGATGCTGAGCNCTAGAGATATGTGCCTGAGCTCC-ACAGTTCTTCAAAGGGAAGGTGAA 69

Qy        915 GGTGGTGAGCCTGAAGCCTGAGGTGGCCCAGGTCGACCTGTACATCCTCGGCCAAGCCGA 974
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         70 GGTGGTGAGCCTGAAGCCTGAGGTGGCCCAGGTCGACCTGTACATCCTCGGCCAAGCCGA 129

Qy        975 CCACTTTATTGGCAACTGTGTCTCCTCCTTCACTGCCTTTGTGAAGCGGGAGCGGGACCT 1034
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        130 CCACTTTATTGGCAACTGTGTCTCCTCCTTCACTGCCTTTGTGAAGCGGGAGCGGGACCT 189

Qy       1035 CCAGGGGAGGCCGTCTTCTTTCTTCGGCATGGACAGGCCCCCTAAGCTGCGGGACGAGTT 1094
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        190 CCAGGGGAGGCCGTCTTCTTTCTTCGGCATGGACAGGCCCCCTAAGCTGCGGGACGAGTT 249

Qy       1095 CTGATT 1100
              ||||||
Db        250 CTGATT 255
```

RESULT 14

```
ABL13947
ID   ABL13947 standard; cDNA; 1209 BP.
XX
AC   ABL13947;
XX
DT   26-MAR-2002  (first entry)
XX
DE   Drosophila melanogaster expressed polynucleotide SEQ ID NO 36323.
XX
KW   Drosophila; developmental biology; cell signalling; insecticide;
KW   pharmaceutical; gene; ss.
XX
OS   Drosophila melanogaster.
XX
PN   WO200171042-A2.
XX
PD   27-SEP-2001.
XX
PF   23-MAR-2001; 2001WO-US009231.
XX
PR   23-MAR-2000; 2000US-0191637P.
PR   11-JUL-2000; 2000US-00614150.
XX
PA   (PEKE ) PE CORP NY.
XX
PI   Venter JC,  Adams M,  Li PWD,  Myers EW;
XX
DR   WPI; 2001-656860/75.
DR   P-PSDB; ABB69844.
XX
PT   New isolated nucleic acid detection reagent for detecting 1000 or more
PT   genes from Drosophila and for elucidating cell signaling and cell-cell
PT   interactions.
XX
PS   Claim 1; SEQ ID NO 36323; 21pp + Sequence Listing; English.
XX
CC   The invention relates to an isolated nucleic acid detection reagent
CC   capable of detecting 1000 or more genes from Drosophila. The invention is
CC   useful in developmental biology and in elucidating cell signalling and
CC   cell-cell interactions in higher eukaryotes for the development of
CC   insecticides, therapeutics and pharmaceutical drugs. The invention
CC   discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC   sequences (ABL01840-ABL16175) and the encoded proteins (ABB57737-
CC   ABB72072). The sequence data for this patent did not form part of the
CC   printed specification, but was obtained in electronic format directly
CC   from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX
SQ   Sequence 1209 BP; 302 A; 313 C; 307 G; 287 T; 0 U; 0 Other;

  Query Match             15.1%;  Score 166.4;  DB 4;  Length 1209;
  Best Local Similarity   51.5%;  Pred. No. 3.7e-37;
  Matches  563;  Conservative    0;  Mismatches  486;  Indels   45;  Gaps    6;

Qy         17 GGGACCCGGCCGGTTACCTGCTCTACTGCCCCTGCATGGGGCGCTTTGGGAACCAGGCCG 76
Db         80 GCGATCCCAATGGCTACCTCACCTACTGTCCGTGTATGGGACGCTTTGGCAACCAGGCCG 139

Qy         77 ATCACTTCTTGGGCTCTCTGGCATTTGCAAAGCTGCTAAACCGTACCTTGGCTGTCCCTC 136
Db        140 ACCACTTCCTGGGATCATTGGCCTTCGCCAAGGCGCTTAATCGCACCCTGATCCTGCCGC 199

Qy        137 CTTGGATTGAGTACCAGCATCACAAGCCTCCTTTCACCAACCTCCATGTGTCCTACCAGA 196
Db        200 CGTGGGTGGAGT------ATCGTAGGGGTGAACTGCGATCCCGTCAGGTACCGTTCAACA 253

Qy        197 AGTACTTCAAGCTGGAGCCCCTCCAGGCTTACCATCGGGTCATCAGCTTGGAGGATTTCA 256
Db        254 CATATTTGAAGTGGAGCCCCTGAAGGAATACCATCGCGTCATCACCATGGCAGATTTCA 313

Qy        257 TGGAGAAGCTGGCACCCACCCACTGGCCCCCTGAGAAGCGGGTGGCATACTGCTTTGAGG 316
Db        314 TGTGGCACCTGGCCGACGACATTTGGCCAGAATCGGAGCGAGTGTCATTTTGCTACAAGG 373

Qy        317 ----------------TGGCAGCCCAGCGAAGCCCAGATAAGAAGACGTGCCCCATGAAGG 361
Db        374 AACGATATAGCCTTCAGCAGGAGAAGAACGATCCAGACAAGCCCAATTGCCACGCCAAGG 433

Qy        362 AAGGAAACCCCTTTGGCCCATTCTGGGATCAGTTTCATGTGAGTTTCAACAAGTCGGAGC 421
Db        434 ATGGCAATCCTTTTGGTCCCTTTTGGGACACTTTTCACATAGACTTTGTGCGGTCAGAGT 493

Qy        422 TTTTTACAGGCATTTCCTTCAGTGCTTCCTACAG---------AGAACAATGGAGCCAGA 472
Db        494 TCTATGCGCCACTTCATTTTGATGTGCATCATAGCAACGAGGCTGCCAAGTGGCAGACCA 553

Qy        473 GATTTTCTCCAAAGGAACATCCGGTGCTTGCCCTGCCAGGAGCCCCAGCCCAGTTCCCCG 532
Db        554 AATATCCTGCAGAATCATATCCCGTACTCGCGTTCACCGGAGCTCCGGCTAGTTTTCCTG 613

Qy        533 TCCTAGAAGAACACAGGCCACTACAGAAGTACATGGTATGGTCAGACGAAATGGTGAAGA 592
Db        614 TTCAGCTAGAGAACTGCAAGCTGCAGCGCTACTTGCAGTGGAGTCAACGGTATAGGGAAG 673

Qy        593 CGGGAGAGGCCCAGATTCATGCCCACCTTGTCCGGC---CCTATGTGGGCATTCATCTGC 649
Db        674 CATCTAAGGATTTCATCCGAGAGCAGTTGCCTCGGGGTGCCTTTTTGGGCATTCATCTGC 733

Qy        650 GCATTGGCTCTGACTGGAAGAACGCCTGTGCCATGCTGAAGGACGG---------GACTG 700
Db        734 GCAACGGTATCGATTGGGTGAGAGCCTGTGAGCACGTCAAGGATAGCCAGCATCTGTTTG 793

Qy        701 CAGGCTCGCACTTCATGGCCTCTCCGCAGTGTGTGGGCTACAGCCGCAGCACAGCGGCCC 760
Db        794 CCTCGCCGCAGTGCTTGGGCTATAAAAATGAACGTGGTGCACTCTACCCGGAGCTCTGCA 853

Qy        761 CCCTCACGATGACTATGTGCCTGCCTGACCTGAAGGAGATCCAGAGGGCTGTGAAGCTCT 820
Db        854 TGCCCTCCAAGGAGGCGATCATCCGCCAGCTAAAGAGAACCATTAAGAACGTGCGCCAAA 913

Qy        821 GGGTGAGGTCGCTGGATGCCCAGTCGGTCTACGTTGCTACTGATTCCGAGAGTTATGTGC 880
Db        914 CTCAGCCGGACAACGAAATCAAATCAGTTTTCGTGGCGTCAGACTCCAATCACATGATTG 973

Qy        881 CTGAGCTCCAACAGCTCTTCAAAGGGAAGGTGAAGGTGGTGAGCCTGAAGCCTGAGGTGG 940
Db        974 GTGAACTAAACACGGCCCTTAGTCGCATGGGCATCAGTGTGCACAAGCTGCCGGAGGATG 1033

Qy        941 ---CCCAGGTCGACCTGTACATCCTCGGCCAAGCCGACCACTTTATTGGCAACTGTGTCT 997
Db       1034 ATCCTTACCTGGACTTGGCCATTCTCGGACAGTCGAACCACTTTATCGGCAACTGTATAT 1093

Qy        998 CCTCCTTCACTGCCTTTGTGAAGCGGGAGCGGGACCTCCAGGGGAGGCCGTCTTCTTTCT 1057
Db       1094 CCTCTTACTCGGCATTCGAAAAAAGGGAACGAGATGTGCACGGTTTTCCATCGTACTTCT 1153

Qy       1058 TCGGCATGGACAGG 1071
Db       1154 GGGGATTCCCCAAG 1167


RESULT 15
ABL13946/c
ID   ABL13946 standard; cDNA; 3264 BP.
XX
AC   ABL13946;
XX
DT   26-MAR-2002  (first entry)
XX
DE   Drosophila melanogaster expressed polynucleotide SEQ ID NO 36320.
XX
KW   Drosophila; developmental biology; cell signalling; insecticide;
KW   pharmaceutical; gene; ss.
XX
OS   Drosophila melanogaster.
XX
PN   WO200171042-A2.
```

XX
PD 27-SEP-2001.
XX
PF 23-MAR-2001; 2001WO-US009231.
XX
PR 23-MAR-2000; 2000US-0191637P.
PR 11-JUL-2000; 2000US-00614150.
XX
PA (PEKE) PE CORP NY.
XX
PI Venter JC, Adams M, Li PWD, Myers EW;
XX
DR WPI; 2001-656860/75.
DR P-PSDB; ABB69843.
XX
PT New isolated nucleic acid detection reagent for detecting 1000 or more
PT genes from Drosophila and for elucidating cell signaling and cell-cell
PT interactions.
XX
PS Claim 1; SEQ ID NO 36320; 21pp + Sequence Listing; English.
XX
CC The invention relates to an isolated nucleic acid detection reagent
CC capable of detecting 1000 or more genes from Drosophila. The invention is
CC useful in developmental biology and in elucidating cell signalling and
CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC sequences (ABL01840-ABL16175) and the encoded proteins (ABB57737-
CC ABB72072). The sequence data for this patent did not form part of the
CC printed specification, but was obtained in electronic format directly
CC from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 3264 BP; 904 A; 742 C; 758 G; 860 T; 0 U; 0 Other;

Query Match 13.7%; Score 150.4; DB 4; Length 3264;
Best Local Similarity 50.9%; Pred. No. 2.4e-32;
Matches 541; Conservative 0; Mismatches 476; Indels 45; Gaps 6;



Search completed: November 9, 2005, 04:43:20
Job time : 644.158 secs

GenCore version 5.1.6

OM nucleic - nucleic search, using sw model

Run on: November 9, 2005, 00:12:44 ; Search time 3988.87 Seconds
(without alignments)
10496.888 Million cell updates/sec

Title: US-09-774-954-16
Perfect score: 1100
Sequence: 1 ATGCCCGCGGGCTCCTGGGA.....CTGCGGGACGAGTTCTGATT 1100

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 34239544 seqs, 19032134700 residues

Total number of hits satisfying chosen parameters: 68479088

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : EST:*
1: gb_est1:*
2: gb_est2:*
3: gb_htc:*
4: gb_est3:*
5: gb_est4:*
6: gb_est5:*
7: gb_est6:*
8: gb_gss1:*
9: gb_gss2:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	1096.8	99.7	1555	3	CR617950	full-leng
2	892.8	81.2	1141	5	BX353274	BX353274 BX353274
3	885.4	80.5	3611	3	AK081059	AK081059 Mus muscu
4	884.2	80.4	2616	3	AK044629	AK044629 Mus muscu
5	743.6	67.6	1091	5	BM909055	BM909055 AGENCOURT
6	718.4	65.3	841	5	BX345989	BX345989 BX345989
7	717.4	65.2	719	1	AL135434	AL135434 DKFZp762O
8	712.2	64.7	889	4	BI524031	BI524031 603052202
9	706.6	64.2	898	5	BX449172	BX449172 BX449172
10	673.2	61.2	942	7	CO774833	ILLUMIGEN
11	662.2	60.2	808	4	BG747760	BG747760 602705255
12	660.2	60.0	801	5	BP144264	BP144264 BP144264
13	639	58.1	988	5	BQ964179	AGENCOURT
14	636.4	57.9	874	4	BI757540	BI757540 603029693
15	630.8	57.3	947	4	BI410783	BI410783 602963556
16	625.6	56.9	809	4	BG473023	BG473023 602514976
17	624	56.7	915	5	BQ936634	BQ936634 AGENCOURT
18	609.8	55.4	705	7	CN348925	CN348925 170005328
19	604.6	55.0	919	5	BU155142	AGENCOURT
20	599.2	54.5	753	4	BG424731	BG424731 602453575
21	578.2	52.6	689	4	BI559635	BI559635 603252494
22	558.4	50.8	561	4	BM798261	BM798261 K-EST0081
23	542.4	49.3	760	5	BU055535	BU055535 UI-M-FO0-
24	541.8	49.3	545	7	CN348928	CN348928 170005313
c 25	514.2	46.7	889	5	BX345988	BX345988 BX345988
26	511.2	46.5	763	7	CK366990	AGENCOURT
27	510	46.4	624	7	CN348922	CN348922 170004249
28	500.8	45.5	666	5	BP459416	BP459416 BP459416
29	499.4	45.4	699	1	AU137536	AU137536 AU137536
30	498	45.3	590	4	BG894786	BG894786 355525 MA
31	497	45.2	592	4	BM714731	BM714731 UI-E-EJ0-
32	497	45.2	608	4	BM711905	BM711905 UI-E-CL1-
33	494.2	44.9	736	5	BP457865	BP457865 BP457865
34	491.8	44.7	769	1	AU117392	AU117392 AU117392
35	485	44.1	644	7	CN459072	CN459072 UI-M-HN0-
36	483.4	43.9	509	2	BE878197	BE878197 601487689
37	471.4	42.9	656	7	CV029252	CV029252 7968 Full
38	471.4	42.9	898	2	BF315757	BF315757 601900828
39	471.4	42.9	955	2	BE260030	BE260030 601150556
40	469.8	42.7	790	4	BG519639	BG519639 602578756
41	469.6	42.7	4147	3	AK053648	AK053648 Mus muscu
42	463.4	42.1	914	5	BQ677046	BQ677046 AGENCOURT
43	457.8	41.6	928	7	CF239348	CF239348 AGENCOURT
44	452.4	41.1	664	6	BY728337	BY728337 BY728337
45	451.4	41.0	661	2	BB642654	BB642654 BB642654

ALIGNMENTS

RESULT 1
CR617950
LOCUS CR617950 1555 bp mRNA linear HTC 21-JUL-2004
DEFINITION full-length cDNA clone CS0DJ008YF12 of T cells (Jurkat cell line) Cot 10-normalized of Homo sapiens (human).
ACCESSION CR617950
VERSION CR617950.1 GI:50498757
KEYWORDS HTC; CNSLT_cDNA.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 1555)
AUTHORS Li,W.B., Gruber,C., Jessee,J. and Polayes,D.
TITLE Full-length cDNA libraries and normalization
JOURNAL Unpublished
REMARK Contact : Feng Liang Email : fliang@lifetech.com URL : http://fulllength.invitrogen.com/ InVitroGen Corporation 1600 Faraday Avenue
REFERENCE 2 (bases 1 to 1555)
AUTHORS Genoscope.
TITLE Direct Submission
JOURNAL Submitted (20-JUL-2004) Genoscope - Centre National de Sequencage : BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr - Web : www.genoscope.cns.fr)
COMMENT 1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime end enriched, double-strand cDNA was digested with Not I and cloned into the Not I and EcoR V sites of the pCMVSPORT 6 vector. Library was normalized. Library was constructed by Life Technologies, a division of Invitrogen.
FEATURES Location/Qualifiers
source 1. .1555
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="CS0DJ008YF12"
/tissue_type="T cells (Jurkat cell line) Cot 10-normalized"
/plasmid="pCMVSPORT_6"
ORIGIN

Query Match 99.7%; Score 1096.8; DB 3; Length 1555;
Best Local Similarity 99.8%; Pred. No. 5e-291;
Matches 1098; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 ATGCCCGCGGGCTCCTGGGACCCGGCCGGTTACCTGCTCTACTGCCCCTGCATGGGGCGC 60
||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||||

```
Db        109 ATGCCTGCGGGCTCCTGGGACCCGGCCGGTTACCTGCTCTACTGCCCCTGCATGGGGCGC 168

Qy         61 TTTGGGAACCAGGCCGATCACTTCTTGGGCTCTCTGGCATTTGCAAAGCTGCTAAACCGT 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        169 TTTGGGAACCAGGCCGATCACTTCTTGGGCTCTCTGGCATTTGCAAAGCTGCTAAACCGT 228

Qy        121 ACCTTGGCTGTCCCTCCTTGGATTGAGTACCAGCATCACAAGCCTCCTTTCACCAACCTC 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        229 ACCTTGGCTGTCCCTCCTTGGATTGAGTACCAGCATCACAAGCCTCCTTTCACCAACCTC 288

Qy        181 CATGTGTCCTACCAGAAGTACTTCAAGCTGGAGCCCCTCCAGGCTTACCATCGGGTCATC 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        289 CATGTGTCCTACCAGAAGTACTTCAAGCTGGAGCCCCTCCAGGCTTACCATCGGGTCATC 348

Qy        241 AGCTTGGAGGATTTCATGGAGAAGCTGGCACCCACCCACTGGCCCCCTGAGAAGCGGGTG 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        349 AGCTTGGAGGATTTCATGGAGAAGCTGGCACCCACCCACTGGCCCCCTGAGAAGCGGGTG 408

Qy        301 GCATACTGCTTTGAGGTGGCAGCCCAGCGAAGCCCAGATAAGAAGACGTGCCCCATGAAG 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        409 GCATACTGCTTTGAGGTGGCAGCCCAGCGAAGCCCAGATAAGAAGACGTGCCCCATGAAG 468

Qy        361 GAAGGAAACCCCTTTGGCCCATTCTGGGATCAGTTTCATGTGAGTTTCAACAAGTCGGAG 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        469 GAAGGAAACCCCTTTGGCCCATTCTGGGATCAGTTTCATGTGAGTTTCAACAAGTCGGAG 528

Qy        421 CTTTTTACAGGCATTTCCTTCAGTGCTTCCTACAGAGAACAATGGAGCCAGAGATTTTCT 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        529 CTTTTTACAGGCATTTCCTTCAGTGCTTCCTACAGAGAACAATGGAGCCAGAGATTTTCT 588

Qy        481 CCAAAGGAACATCCGGTGCTTGCCCTGCCAGGAGCCCCAGCCCAGTTCCCCGTCCTAGAA 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        589 CCAAAGGAACATCCGGTGCTTGCCCTGCCAGGAGCCCCAGCCCAGTTCCCCGTCCTAGAG 648

Qy        541 GAACACAGGCCACTACAGAAGTACATGGTATGGTCAGACGAAATGGTGAAGACGGGAGAG 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        649 GAACACAGGCCACTACAGAAGTACATGGTATGGTCAGACGAAATGGTGAAGACGGGAGAG 708

Qy        601 GCCCAGATTCATGCCCACCTTGTCCGGCCCTATGTGGGCATTCATCTGCGCATTGGCTCT 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        709 GCCCAGATTCATGCCCACCTTGTCCGGCCCTATGTGGGCATTCATCTGCGCATTGGCTCT 768

Qy        661 GACTGGAAGAACGCCTGTGCCATGCTGAAGGACGGGACTGCAGGCTCGCACTTCATGGCC 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        769 GACTGGAAGAACGCCTGTGCCATGCTGAAGGACGGGACTGCAGGCTCGCACTTCATGGCC 828

Qy        721 TCTCCGCAGTGTGTGGGCTACAGCCGCAGCACAGCGGCCCCCCTCACGATGACTATGTGC 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        829 TCTCCGCAGTGTGTGGGCTACAGCCGCAGCACAGCGGCCCCCCTCACGATGACTATGTGC 888

Qy        781 CTGCCTGACCTGAAGGAGATCCAGAGGGCTGTGAAGCTCTGGGTGAGGTCGCTGGATGCC 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        889 CTGCCTGACCTGAAGGAGATCCAGAGGGCTGTGAAGCTCTGGGTGAGGTCGCTGGATGCC 948

Qy        841 CAGTCGGTCTACGTTGCTACTGATTCCGAGAGTTATGTGCCTGAGCTCCAACAGCTCTTC 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        949 CAGTCGGTCTACGTTGCTACTGATTCCGAGAGTTATGTGCCTGAGCTCCAACAGCTCTTC 1008

Qy        901 AAAGGGAAGGTGAAGGTGGTGAGCCTGAAGCCTGAGGTGGCCCAGGTCGACCTGTACATC 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1009 AAAGGGAAGGTGAAGGTGGTGAGCCTGAAGCCTGAGGTGGCCCAGGTCGACCTGTACATC 1068

Qy        961 CTCGGCCAAGCCGACCACTTTATTGGCAACTGTGTCTCCTCCTTCACTGCCTTTGTGAAG 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1069 CTCGGCCAAGCCGACCACTTTATTGGCAACTGTGTCTCCTCCTTCACTGCCTTTGTGAAG 1128

Qy       1021 CGGGAGCGGGACCTCCAGGGGAGGCCGTCTTCTTTCTTCGGCATGGACAGGCCCCCTAAG 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1129 CGGGAGCGGGACCTCCAGGGGAGGCCGTCTTCTTTCTTCGGCATGGACAGGCCCCCTAAG 1188

Qy       1081 CTGCGGGACGAGTTCTGATT 1100
              ||||||||||||||||||||
Db       1189 CTGCGGGACGAGTTCTGATT 1208
```

```
RESULT 2
BX353274
LOCUS       BX353274               1141 bp    mRNA    linear   EST 08-APR-2004
DEFINITION  BX353274 Homo sapiens NEUROBLASTOMA COT 25-NORMALIZED Homo sapiens
            cDNA clone CS0DC003YJ11 5-PRIME, mRNA sequence.
ACCESSION   BX353274
VERSION     BX353274.2  GI:46290116
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1141)
  AUTHORS   Li,W.B., Gruber,C., Jessee,J. and Polayes,D.
  TITLE     Full-length cDNA libraries and normalization
  JOURNAL   Unpublished (2001)
COMMENT     On May 5, 2003 this sequence version replaced gi:30371746.
            Contact: Genoscope
            Genoscope - Centre National de Sequencage
            2 rue Gaston Cremieux, CP 5706 - 91057 EVRY cedex - FRANCE
            Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr
            1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime
            end enriched, double-strand cDNA was digested with Not I and cloned
            into the Not I and EcoR V sites of the pCMVSPORT 6 vector. Library
            was normalized. Library was constructed by Life Technologies, a
            division of Invitrogen. This sequence belongs to sequence cluster
            9980.f
            For more information about this cluster, see
            http://www.genoscope.cns.fr/cdna?s=CS0DC003CE06QP1&c=9980.f.
FEATURES             Location/Qualifiers
     source          1. .1141
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="CS0DC003YJ11"
                     /tissue_type="NEUROBLASTOMA COT 25-NORMALIZED"
                     /clone_lib="Homo sapiens NEUROBLASTOMA COT 25-NORMALIZED"
                     /note="1st strand cDNA was primed with a NotI-oligo(dT)
                     primer. Five prime end enriched, double-strand cDNA was
                     digested with Not I and cloned into the Not I and EcoR V
                     sites of the pCMVSPORT 6 vector. Library was normalized."
ORIGIN
```

```
  Query Match             81.2%;  Score 892.8;  DB 5;  Length 1141;
  Best Local Similarity   92.8%;  Pred. No. 8.9e-235;
  Matches  952;  Conservative   32;  Mismatches   35;  Indels    7;  Gaps    5;

Qy          1 ATGCCCGCGGGCTCCTGGGACCCGGCCGGTTACCTGCTCTACTGCCCCTGCATGGGGCGC 60
              ||||||||:||||||||||||||||||||||||||||||||||||:||||||||||||||
Db         86 ATGCCCGCRGGCTCCTGGGACCCGGCCGGTTACCTGCTCTACTGCMCCTGCATGGGGCGC 145

Qy         61 TTTGGGAACCAGGCCGATCACTTCTTGGGCTCTCTGGCATTTGCAAAGCTGCTAAACCGT 120
              ||||||||:||||||||:|||||||||||||||||||||||||||||||||||:|||::||
Db        146 TTTGGGAAMCAGGCCGATMACTTCTTGGGCTCTCTGGCATTTGCAAAGCTGCWAAAMMGT 205

Qy        121 ACCTTGGCTGTCCCTCCTTGGATTGAGTACCAGCATCACAAGCCTCCTTTCACCAACCTC 180
              |||||||||| ||||||||||||||||||||||||:||||:|||||||||||||||||||
Db        206 ACCTTGGCTGTACCTCCTTGGATTGAGTACCAGCATCAMAAGCMTCCTTTCACCAACCTC 265

Qy        181 CATGTGTCCTACCAGAAGTACTTCAAGCTGGAGCCCCTCCAGGCTTACCATCGGGTCATC 240
              ||||||||||||||||||||||:||:::|||||  |||:||||||||:||||||||||||
Db        266 CATGTGTCCTACCAGAAGTACTTMAARMWGGAGCACCTMCAGGCTTACMATCGGGTCATC 325

Qy        241 AGCTTGGAGGATTTCATGGAGAAGCTGGCACCCACCCACTGGCCCCCTGAGAAGCGGGTG 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        326 AGCTTGGAGGATTTCATGGAGAAGCTGGCACCCACCCACTGGCCCCCTGAGAAGCGGGTG 385

Qy        301 GCATACTGCTTTGAGGTGGCAGCCCAGCGAAGCCCAGATAAGAAGACGTGCCCCATGAAG 360
              ||||||||||||||||||||||||:|||||||| |||||:||||||||||||||||||||
Db        386 GCATACTGCTTTGAGGTGGCAGCMCAGCGAAG-CCAGAWAAGAAGACGTGCCCCATGAAG 444
```

```
Qy        361 GAAGGAAACCCCTTTGGCCCATTCTGGGATCAGTTTCATGTGAGTTTCAACAAGTCGGAG 420
              ||||||||||||||||||||||||||||||||||||||||||||| ||||||||||||||
Db        445 GAAGGAAACCCCTTTGGCCCATTCTGGGATCAGTTTCATGTGAGTATCAACAAGTCGGAG 504

Qy        421 CTTTTTACAGGCATTTCCTTCAGTGCTTCCTACAGAGAACAATGGAGCCAGAGATTTTCT 480
              ||||||||||||||||:|||||||||||||||||||||||||||||||||||||||||||
Db        505 CTTTTTACAGGCATTTMCTTCAGTGCTTCCTACAGAGAACAATGGAGCCAGAGATTTTCT 564

Qy        481 CCAAAGGAACATCCGGTGCTTGCCCTGCCAGGAGCCCCAGCCCAGTTCCCCGTCCTAGAA 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        565 CCAAAGGAACATCCGGTGCTTGCCCTGCCAGGAGCCCCAGCCCAGTTCCCCGTCCTAGAG 624

Qy        541 GAACACAGGCCACTACAGAAGTACATGGTATGGTCAGACGAAATGGTGAAGACGGGAGAG 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||:
Db        625 GAACACAGGCCACTACAGAAGTACATGGTATGGTCAGACGAAATGGTGAAGACGGGAGAR 684

Qy        601 GCCCAGATTCATGCCCACCTTGTCCGGCCCTATGTGGGCATTCATCTGCGCATTGGCTCT 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        685 GCCCAGATTCATGCCCACCTTGTCCGGCCCTATGTGGGCATTCATCTGCGCATTGGCTCT 744

Qy        661 GACTGGAAGAACGCCTGTGCCATGCTGAAGGACGGGACTGCAGGCTCGCACTTCATGGCC 720
              ||||||||||||||||:|||||||||||||||||||||||||||||||||||||||||||
Db        745 GACTGGAAGAACGCCTGTDCCATGCTGAAGGACGGGACTGCAGGCTCGCACTTCATGGCC 804

Qy        721 TCTCCGCAGTGTGTGGGCTACAGCCGCAGCACAGCGGCCCCCCTCACGATGACTATGTGC 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        805 TCTCCGCAGTGTGTGGGCTACAGCCGCAGCACAGCGGCCCCCCTCACGATGACTATGTGC 864

Qy        781 CTGCCTGACCTGAAGGAGATCCAGAGGGCTGTGAAGCTCT-GGGTGAGGTCGCTGGATGC 839
              ||||||||||||||||||||||||||||||:|||||||| ||||||||||||||||||||
Db        865 CTGCCTGACCTGAAGGAGATCCAGAGGGCTGTKAAGCTCTGGGGTGAGGTCGCTGGATGC 924

Qy        840 CCAGTCGGTCTACGTTGCTACTGATTCCGAGAGTTATGTGCCTGAGCTCCAACAGCTCTT 899
              |||||||||||||||||||||||||||||||||||| |||:|||||||||||||||||
Db        925 CCAGTCGGTCTACGTTGCTACTGATTCCGAGAGTTATGTTCCTKAGCTCCAACAGCTCTT 984

Qy        900 CAAAGGGAAGGTGAAGGTGGTGAGCCTGAAGCCTGAGGTGGCCCAGGTCGACCTGTACAT 959
              |||||||||||||||||||||||||| ||:||||:|||||:|:||   |     |||||:||
Db        985 CAAAGGGAAGGTGAAGGTGGTGAG-CTKAAGCYTGAGGKGSCC--AGGTCGACTGTAMAT 1041

Qy        960 CCTCGGCCAAGCCGACCACTTTATTGGCAACTGTGTCTCCTCCTTCACTGCCTTTGTGAA 1019
              |||||||  |  ||: |||||||||||||| : : |   |:   |  ||   |:   |
Db       1042 CCTCGGC--AAGCGRACACTTTATTGGCAATKTKTTTCTCYTTATTGCTTTGTKAAGGGG 1099

Qy       1020 GCGGGA 1025
                 ||||
Db       1100 AGGGGA 1105
```

```
RESULT 3
AK081059
LOCUS       AK081059               3611 bp    mRNA    linear   HTC 03-APR-2004
DEFINITION  Mus musculus 10 days neonate cerebellum cDNA, RIKEN full-length
            enriched library, clone:B930076G19 product:protein
            O-fucosyltransferase 1, full insert sequence.
ACCESSION   AK081059
VERSION     AK081059.1  GI:26099647
KEYWORDS    HTC; CAP trapper.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1
  AUTHORS   Carninci,P. and Hayashizaki,Y.
  TITLE     High-efficiency full-length cDNA cloning
  JOURNAL   Meth. Enzymol. 303, 19-44 (1999)
  MEDLINE   99279253
   PUBMED   10349636
REFERENCE   2
  AUTHORS   Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K.,
            Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
  TITLE     Normalization and subtraction of cap-trapper-selected cDNAs to
            prepare full-length cDNA libraries for rapid discovery of new genes
  JOURNAL   Genome Res. 10 (10), 1617-1630 (2000)
  MEDLINE   20499374
   PUBMED   11042159
REFERENCE   3
  AUTHORS   Shibata,K., Itoh,M., Aizawa,K., Nagaoka,S., Sasaki,N., Carninci,P.,
            Konno,H., Akiyama,J., Nishi,K., Kitsunai,T., Tashiro,H., Itoh,M.,
            Sumi,N., Ishii,Y., Nakamura,S., Hazama,M., Nishine,T., Harada,A.,
            Yamamoto,R., Matsumoto,H., Sakaguchi,S., Ikegami,T., Kashiwagi,K.,
            Fujiwake,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E., Watahiki,M.,
            Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsuura,S., Kawai,J.,
            Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and Hayashizaki,Y.
  TITLE     RIKEN integrated sequence analysis (RISA) system--384-format
            sequencing pipeline with 384 multicapillary sequencer
  JOURNAL   Genome Res. 10 (11), 1757-1771 (2000)
  MEDLINE   20530913
   PUBMED   11076861
REFERENCE   4
  AUTHORS   The RIKEN Genome Exploration Research Group Phase II Team and the
            FANTOM Consortium.
  TITLE     Functional annotation of a full-length mouse cDNA collection
  JOURNAL   Nature 409, 685-690 (2001)
REFERENCE   5
  AUTHORS   The FANTOM Consortium and the RIKEN Genome Exploration Research
            Group Phase I & II Team.
  TITLE     Analysis of the mouse transcriptome based on functional annotation
            of 60,770 full-length cDNAs
  JOURNAL   Nature 420, 563-573 (2002)
REFERENCE   6  (bases 1 to 3611)
  AUTHORS   Adachi,J., Aizawa,K., Akimura,T., Arakawa,T., Bono,H., Carninci,P.,
            Fukuda,S., Furuno,M., Hanagaki,T., Hara,A., Hashizume,W.,
            Hayashida,K., Hayatsu,N., Hiramoto,K., Hiraoka,T., Hirozane,T.,
            Hori,F., Imotani,K., Ishii,Y., Itoh,M., Kagawa,I., Kasukawa,T.,
            Katoh,H., Kawai,J., Kojima,Y., Kondo,S., Konno,H., Kouda,M.,
            Koya,S., Kurihara,C., Matsuyama,T., Miyazaki,A., Murata,M.,
            Nakamura,M., Nishi,K., Nomura,K., Numazaki,R., Ohno,M., Ohsato,N.,
            Okazaki,Y., Saito,R., Saitoh,H., Sakai,C., Sakai,K., Sakazume,N.,
            Sano,H., Sasaki,D., Shibata,K., Shinagawa,A., Shiraki,T.,
            Sogabe,Y., Tagami,M., Tagawa,A., Takahashi,F., Takaku-Akahira,S.,
            Takeda,Y., Tanaka,T., Tomaru,A., Toya,T., Yasunishi,A.,
            Muramatsu,M. and Hayashizaki,Y.
  TITLE     Direct Submission
  JOURNAL   Submitted (16-APR-2002) Yoshihide Hayashizaki, The Institute of
            Physical and Chemical Research (RIKEN), Laboratory for Genome
            Exploration Research Group, RIKEN Genomic Sciences Center (GSC),
            RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama,
            Kanagawa 230-0045, Japan (E-mail:genome-res@gsc.riken.jp,
            URL:http://genome.gsc.riken.jp/, Tel:81-45-503-9222,
            Fax:81-45-503-9216)
COMMENT     cDNA library was prepared and sequenced in Mouse Genome
            Encyclopedia Project of Genome Exploration Research Group in Riken
            Genomic Sciences Center and Genome Science Laboratory in RIKEN.
            Division of Experimental Animal Research in Riken contributed to
            prepare mouse tissues.
            Please visit our web site for further details.
            URL:http://genome.gsc.riken.jp/
            URL:http://fantom.gsc.riken.jp/.
FEATURES             Location/Qualifiers
     source          1. .3611
                     /organism="Mus musculus"
                     /mol_type="mRNA"
                     /strain="C57BL/6J"
                     /db_xref="FANTOM_DB:B930076G19"
                     /db_xref="taxon:10090"
                     /clone="B930076G19"
                     /tissue_type="cerebellum"
                     /clone_lib="RIKEN full-length enriched mouse cDNA library"
                     /dev_stage="10 days neonate"
     misc_feature    11. .1194
                     /note="protein O-fucosyltransferase 1 (MGD|MGI:2153207,
                     GB|NM_080463, evidence: BLASTN, 99%, match=1290)
                     putative"
ORIGIN
```

```
 Query Match             80.5%;  Score 885.4;  DB 3;  Length 3611;
 Best Local Similarity   88.0%;  Pred. No. 1.3e-232;
 Matches  964;  Conservative    0;  Mismatches  131;  Indels    0;  Gaps    0;

Qy          5 CCGCGGGCTCCTGGGACCCGGCCGGTTACCTGCTCTACTGCCCCTGCATGGGGCGCTTTG 64
              | ||||||||||||||| |||||||||||||||||||| |||||||||||||||||||
Db        101 CAGCGGGCTCCTGGGACCTGGCCGGTTACCTGCTCTACTGTCCCTGCATGGGGCGCTTTG 160

Qy         65 GGAACCAGGCCGATCACTTCTTGGGCTCTCTGGCATTTGCAAAGCTGCTAAACCGTACCT 124
              |||||||||| |||||||||||||||||| |||||||||||| |||||||| ||||| ||||
Db        161 GGAACCAGGCTGATCACTTCTTGGGCTCCCTGGCATTTGCGAAGCTGCTGAACCGCACCT 220

Qy        125 TGGCTGTCCCTCCTTGGATTGAGTACCAGCATCACAAGCCTCCTTTCACCAACCTCCATG 184
              ||||||| ||||| |||||||| ||||| ||||||||||||||||||||||||||||||
Db        221 TGGCTGTACCTCCATGGATTGAATACCAACATCACAAGCCTCCTTTCACCAACCTCCATG 280

Qy        185 TGTCCTACCAGAAGTACTTCAAGCTGGAGCCCCTCCAGGCTTACCATCGGGTCATCAGCT 244
              |||||||||| |||||||||||| |||||||| ||||| || |||||||||||  |||||
Db        281 TGTCCTACCAAAAGTACTTCAAACTGGAGCCTCTCCAAGCCTACCATCGGGTTGTCAGCC 340

Qy        245 TGGAGGATTTCATGGAGAAGCTGGCACCCACCCACTGGCCCCCTGAGAAGCGGGTGGCAT 304
              ||||||| |||||||| || |||||||| ||||||||||||||||||||||||||||||
Db        341 TGGAGGACTTCATGGAAAATCTGGCACCCTCCCACTGGCCCCCTGAGAAGCGAGTGGCAT 400

Qy        305 ACTGCTTTGAGGTGGCAGCCCAGCGAAGCCCAGATAAGAAGACGTGCCCCATGAAGGAAG 364
              |||||||||||||||||||||||||||| || |||||||||||| || ||||||||||||
Db        401 ACTGCTTTGAGGTGGCAGCCCAGCGAAGTCCTGATAAGAAGACATGTCCCATGAAGGAAG 460

Qy        365 GAAACCCCTTTGGCCCATTCTGGGATCAGTTTCATGTGAGTTTCAACAAGTCGGAGCTTT 424
              |||| || ||||| |||||||||| |||||||||||||||||||||||||||| || || |
Db        461 GAAATCCTTTTGGGCCATTCTGGGACCAGTTTCATGTGAGTTTCAATAAGTCAGAACTGT 520

Qy        425 TTACAGGCATTTCCTTCAGTGCTTCCTACAGAGAACAATGGAGCCAGAGATTTTCTCCAA 484
              | |||||||||||||||| || ||||||| |||||||||| |||||||||| || |||
Db        521 TCACAGGCATTTCCTTCAGCGCCTCCTACAAAGAACAATGGACCCAGAGATTTCCTGCAA 580

Qy        485 AGGAACATCCGGTGCTTGCCCTGCCAGGAGCCCCAGCCCAGTTCCCCGTCCTAGAAGAAC 544
              | || ||||| ||||| || ||||| || |||||||| |||||||| ||||| || ||||
Db        581 AAGAGCATCCTGTGCTCGCACTGCCTGGGGCCCCAGCACAGTTCCCTGTCCTGGAGGAAC 640

Qy        545 ACAGGCCACTACAGAAGTACATGGTATGGTCAGACGAAATGGTGAAGACGGGAGAGGCCC 604
              |||||   || |||||||||||||| |||||||| || ||||||| ||||||||||||||
Db        641 ACAGGGAGCTCCAGAAGTACATGGTGTGGTCAGATGAGATGGTGAGGACGGGAGAGGCCC 700

Qy        605 AGATTCATGCCCACCTTGTCCGGCCCTATGTGGGCATTCATCTGCGCATTGGCTCTGACT 664
              |||   ||||||||| ||||||||||||||||||||||||||||||||||||||| ||||
Db        701 TGATCAGTGCCCACCTCGTCCGGCCCTATGTGGGCATTCATCTGCGCATTGGCTCCGACT 760

Qy        665 GGAAGAACGCCTGTGCCATGCTGAAGGACGGGACTGCAGGCTCGCACTTCATGGCCTCTC 724
              ||||||| |||||||||||||||||||| || |||||||| || |||||||||||  || |
Db        761 GGAAGAATGCCTGTGCCATGCTGAAGGATGGAACTGCAGGGTCACACTTCATGGCTTCCC 820

Qy        725 CGCAGTGTGTGGGCTACAGCCGCAGCACAGCGGCCCCCCTCACGATGACTATGTGCCTGC 784
              | |||||||||||||| |||||||||||||| |||| ||||| ||||| |||||||| |
Db        821 CTCAGTGTGTGGGCTATAGCCGCAGCACAGCGACCCCTCTCACCATGACCATGTGCCTCC 880

Qy        785 CTGACCTGAAGGAGATCCAGAGGGCTGTGAAGCTCTGGGTGAGGTCGCTGGATGCCCAGT 844
              |||||||||||||  ||||| |||||||||| ||| |||||||  | ||| |||||   |
Db        881 CTGACCTGAAGGAAATCCAGCGGGCTGTGACGCTTTGGGTGAGAGCACTGAATGCCAGAT 940

Qy        845 CGGTCTACGTTGCTACTGATTCCGAGAGTTATGTGCCTGAGCTCCAACAGCTCTTCAAAG 904
              |||||||| | || || || || ||||| || ||| | ||| |||| ||||||||||||||
Db        941 CGGTCTACATCGCCACAGACTCTGAGAGCTACGTGTCAGAGATCCAGCAGCTCTTCAAAG 1000

Qy        905 GGAAGGTGAAGGTGGTGAGCCTGAAGCCTGAGGTGGCCCAGGTCGACCTGTACATCCTCG 964
              |||||||| |||||||||||||||| || |||||||||||| ||||||||||||||||||
Db       1001 ACAAGGTGAGGGTGGTGAGCCTGAAACCCGAGGTGGCCCAGATCGACCTGTACATCCTCG 1060

Qy        965 GCCAAGCCGACCACTTTATTGGCAACTGTGTCTCCTCCTTCACTGCCTTTGTGAAGCGGG 1024
              |||| || |||||||| ||||| |||||||||||||| |||||||||| ||||||||||
Db       1061 GCCAGGCTGACCACTTCATTGGAAACTGTGTCTCCTCGTTCACTGCCTTCGTGAAGCGGG 1120

Qy       1025 AGCGGGACCTCCAGGGGAGGCCGTCTTCTTTCTTCGGCATGGACAGGCCCCCTAAGCTGC 1084
              |||||||||||||| ||||||| ||| || ||||| |||||||||| ||| |  |||| |
Db       1121 AGCGGGACCTCCATGGGAGGCAGTCGTCCTTCTTTGGCATGGACAGACCCTCCCAGCTTC 1180

Qy       1085 GGGACGAGTTCTGAT 1099
              |||| || || ||||
Db       1181 GGGATGAATTTTGAT 1195
```

RESULT 4
AK044629

LOCUS	AK044629 2616 bp mRNA linear HTC 03-APR-2004
DEFINITION	Mus musculus adult retina cDNA, RIKEN full-length enriched library, clone:A930028F21 product:protein O-fucosyltransferase 1, full insert sequence.
ACCESSION	AK044629
VERSION	AK044629.1 GI:26336652
KEYWORDS	HTC; CAP trapper.
SOURCE	Mus musculus (house mouse)
ORGANISM	Mus musculus Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE	1
AUTHORS	Carninci,P. and Hayashizaki,Y.
TITLE	High-efficiency full-length cDNA cloning
JOURNAL	Meth. Enzymol. 303, 19-44 (1999)
MEDLINE	99279253
PUBMED	10349636
REFERENCE	2
AUTHORS	Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K., Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
TITLE	Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes
JOURNAL	Genome Res. 10 (10), 1617-1630 (2000)
MEDLINE	20499374
PUBMED	11042159
REFERENCE	3
AUTHORS	Shibata,K., Itoh,M., Aizawa,K., Nagaoka,S., Sasaki,N., Carninci,P., Konno,H., Akiyama,J., Nishi,K., Kitsunai,T., Tashiro,H., Itoh,M., Sumi,N., Ishii,Y., Nakamura,S., Hazama,M., Nishine,T., Harada,A., Yamamoto,R., Matsumoto,H., Sakaguchi,S., Ikegami,T., Kashiwagi,K., Fujiwake,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E., Watahiki,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsuura,S., Kawai,J., Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and Hayashizaki,Y.
TITLE	RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multicapillary sequencer
JOURNAL	Genome Res. 10 (11), 1757-1771 (2000)
MEDLINE	20530913
PUBMED	11076861
REFERENCE	4
AUTHORS	The RIKEN Genome Exploration Research Group Phase II Team and the FANTOM Consortium.
TITLE	Functional annotation of a full-length mouse cDNA collection
JOURNAL	Nature 409, 685-690 (2001)
REFERENCE	5
AUTHORS	The FANTOM Consortium and the RIKEN Genome Exploration Research Group Phase I & II Team.
TITLE	Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs
JOURNAL	Nature 420, 563-573 (2002)
REFERENCE	6 (bases 1 to 2616)
AUTHORS	Adachi,J., Aizawa,K., Akimura,T., Arakawa,T., Bono,H., Carninci,P., Fukuda,S., Furuno,M., Hanagaki,T., Hara,A., Hashizume,W., Hayashida,K., Hayatsu,N., Hiramoto,K., Hiraoka,T., Hirozane,T., Hori,F., Imotani,K., Ishii,Y., Itoh,M., Kagawa,I., Kasukawa,T., Katoh,H., Kawai,J., Kojima,Y., Kondo,S., Konno,H., Kouda,M., Koya,S., Kurihara,C., Matsuyama,T., Miyazaki,A., Murata,M., Nakamura,M., Nishi,K., Nomura,K., Numazaki,R., Ohno,M., Ohsato,N., Okazaki,Y., Saito,R., Saitoh,H., Sakai,C., Sakai,K., Sakazume,N., Sano,H., Sasaki,D., Shibata,K., Shinagawa,A., Shiraki,T., Sogabe,Y., Tagami,M., Tagawa,A., Takahashi,F., Takaku-Akahira,S., Takeda,Y., Tanaka,T., Tomaru,A., Toya,T., Yasunishi,A., Muramatsu,M. and Hayashizaki,Y.
TITLE	Direct Submission

```
JOURNAL    Submitted (16-JUL-2001) Yoshihide Hayashizaki, The Institute of
           Physical and Chemical Research (RIKEN), Laboratory for Genome
           Exploration Research Group, RIKEN Genomic Sciences Center (GSC),
           RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama,
           Kanagawa 230-0045, Japan (E-mail:genome-res@gsc.riken.jp,
           URL:http://genome.gsc.riken.jp/, Tel:81-45-503-9222,
           Fax:81-45-503-9216)
COMMENT    cDNA library was prepared and sequenced in Mouse Genome
           Encyclopedia Project of Genome Exploration Research Group in Riken
           Genomic Sciences Center and Genome Science Laboratory in RIKEN.
           Division of Experimental Animal Research in Riken contributed to
           prepare mouse tissues.
           Retina RNA was provided by Dr. Stefano Gustincich (Department of
           Neurobiology, Harvard Medical School, 220 Longwood Ave., Boston, MA
           02115, USA) whose assistance is gratefully acknowledged. Please
           visit our web site for further details.
           URL:http://genome.gsc.riken.jp/
           URL:http://fantom.gsc.riken.jp/.
FEATURES             Location/Qualifiers
     source          1. .2616
                     /organism="Mus musculus"
                     /mol_type="mRNA"
                     /strain="C57BL/6J"
                     /db_xref="FANTOM_DB:A930028F21"
                     /db_xref="taxon:10090"
                     /clone="A930028F21"
                     /tissue_type="retina"
                     /clone_lib="RIKEN full-length enriched mouse cDNA library"
                     /dev_stage="adult"
     CDS             34. .1215
                     /note="unnamed protein product; protein
                     O-fucosyltransferase 1 (MGD|MGI:2153207, GB|NM_080463,
                     evidence: BLASTN, 99%, match=1290)
                     putative"
                     /codon_start=1
                     /protein_id="BAC32009.1"
                     /db_xref="GI:26336653"
                     /translation="MGAAAWAPPHLLLRASFLLLLLLLPLRGRSAGSWDLAGYLLYCP
                     CMGRFGNQADHFLGSLAFAKLLNRTLAVPPWIEYQHHKPPFTNLHVSYQKYFKLEPLQ
                     AYHRVVSLEDFMENLAPSHWPPEKRVAYCFEVAAQRSPDKKTCPMKEGNPFGPFWDQF
                     HVSFNKSELFTGISFSASYKEQWTQRFPAKEHPVLALPGAPAQFPVLEEHRELQKYMV
                     WSDEMVRTGEALISSHLVRPYVGIHLRIGSDWKNACAMLKDGTAGSHFMASPQCVGYS
                     RSTATPLTMTMCLPDLKEIQRAVTLWVRALNARSVYIATDSESYVSEIQQLFKDKVRV
                     VSLKPEVAQIDLYILGQADHFIGNCVSSFTAFVKRERDLHGRQSSFFGMDRPSQLRDE
                     F"
ORIGIN

  Query Match             80.4%;  Score 884.2;  DB 3;  Length 2616;
  Best Local Similarity   87.9%;  Pred. No. 2.6e-232;
  Matches  964;  Conservative    0;  Mismatches  133;  Indels    0;  Gaps    0;

Qy          3 GCCCGCGGGCTCCTGGGACCCGGCCGGTTACCTGCTCTACTGCCCCTGCATGGGGCGCTT 62
              | | ||||||||||||||||| |||||||||||||||||||| ||||||||||||||||
Db        120 GTCAGCGGGCTCCTGGGACCTGGCCGGTTACCTGCTCTACTGTCCCTGCATGGGGCGCTT 179

Qy         63 TGGGAACCAGGCCGATCACTTCTTGGGCTCTCTGGCATTTGCAAAGCTGCTAAACCGTAC 122
              |||||||||||| |||||||||||||||| |||||||||| |||||||| ||||| ||
Db        180 TGGGAACCAGGCTGATCACTTCTTGGGCTCCCTGGCATTTGCGAAGCTGCTGAACCGCAC 239

Qy        123 CTTGGCTGTCCCTCCTTGGATTGAGTACCAGCATCACAAGCCTCCTTTCACCAACCTCCA 182
              ||||||||| ||||| |||||||| ||||| ||||||||||||||||||||||||||||
Db        240 CTTGGCTGTACCTCCATGGATTGAATACCAACATCACAAGCCTCCTTTCACCAACCTCCA 299

Qy        183 TGTGTCCTACCAGAAGTACTTCAAGCTGGAGCCCCTCCAGGCTTACCATCGGGTCATCAG 242
              |||||||||||| |||||||||| |||||||| ||||| || |||||||||||  ||||
Db        300 TGTGTCCTACCAAAAGTACTTCAAACTGGAGCCTCTCCAAGCCTACCATCGGGTTGTCAG 359

Qy        243 CTTGGAGGATTTCATGGAGAAGCTGGCACCCACCCACTGGCCCCCTGAGAAGCGGGTGGC 302
              | ||||||| |||||||| || ||||||||| ||||||||||||||||||||||| |||||
Db        360 CCTGGAGGACTTCATGGAAAATCTGGCACCCTCCCACTGGCCCCCTGAGAAGCGAGTGGC 419

Qy        303 ATACTGCTTTGAGGTGGCAGCCCAGCGAAGCCCAGATAAGAAGACGTGCCCCATGAAGGA 362
              |||||||||||||||||||||||||||||| || |||||||||| || ||||||||||

Db        420 ATACTGCTTTGAGGTGGCAGCCCAGCGAAGTCCTGATAAGAAGACATGTCCCATGAAGGA 479

Qy        363 AGGAAACCCCTTTGGCCCATTCTGGGATCAGTTTCATGTGAGTTTCAACAAGTCGGAGCT 422
              |||||| || ||||| |||||||||| |||||||||||||||||||| ||||| || ||
Db        480 AGGAAATCCTTTTGGGCCATTCTGGGACCAGTTTCATGTGAGTTTCAATAAGTCAGAACT 539

Qy        423 TTTTACAGGCATTTCCTTCAGTGCTTCCTACAGAGAACAATGGAGCCAGAGATTTTCTCC 482
              || |||||||||||||||| || ||||||| |||||||||| |||||||||| || |
Db        540 GTTCACAGGCATTTCCTTCAGCGCCTCCTACAAAGAACAATGGACCCAGAGATTTCCTGC 599

Qy        483 AAAGGAACATCCGGTGCTTGCCCTGCCAGGAGCCCCAGCCCAGTTCCCCGTCCTAGAAGA 542
              ||| || ||||| ||||| || ||||| || |||||||| |||||||| ||||| || ||
Db        600 AAAAGAGCATCCTGTGCTCGCACTGCCTGGGGCCCCAGCACAGTTCCCTGTCCTGGAGGA 659

Qy        543 ACACAGGCCACTACAGAAGTACATGGTATGGTCAGACGAAATGGTGAAGACGGGAGAGGC 602
              |||||||   || |||||||||||||| |||||||| || ||||||| ||||||||||||
Db        660 ACACAGGGAGCTCCAGAAGTACATGGTGTGGTCAGATGAGATGGTGAGGACGGGAGAGGC 719

Qy        603 CCAGATTCATGCCCACCTTGTCCGGCCCTATGTGGGCATTCATCTGCGCATTGGCTCTGA 662
              || |||   | ||||||| |||||||||||||||||||||||||||||||||||||| ||
Db        720 CCTGATCAGTTCCCACCTCGTCCGGCCCTATGTGGGCATTCATCTGCGCATTGGCTCCGA 779

Qy        663 CTGGAAGAACGCCTGTGCCATGCTGAAGGACGGGACTGCAGGCTCGCACTTCATGGCCTC 722
              ||||||||| |||||||||||||||||| || |||||||| || |||||||||||| ||
Db        780 CTGGAAGAATGCCTGTGCCATGCTGAAGGATGGAACTGCAGGGTCACACTTCATGGCTTC 839

Qy        723 TCCGCAGTGTGTGGGCTACAGCCGCAGCACAGCGGCCCCCCTCACGATGACTATGTGCCT 782
              || |||||||||||||| |||||||||||||| |||| |||| |||| ||||||||
Db        840 CCCTCAGTGTGTGGGCTATAGCCGCAGCACAGCGACCCCTCTCACCATGACCATGTGCCT 899

Qy        783 GCCTGACCTGAAGGAGATCCAGAGGGCTGTGAAGCTCTGGGTGAGGTCGCTGGATGCCCA 842
              |||||||||||||| |||||| |||||||| ||| |||||||  | ||| ||||
Db        900 CCCTGACCTGAAGGAAATCCAGCGGGCTGTGACGCTTTGGGTGAGAGCACTGAATGCCAG 959

Qy        843 GTCGGTCTACGTTGCTACTGATTCCGAGAGTTATGTGCCTGAGCTCCAACAGCTCTTCAA 902
              |||||||| | || || || || ||||| || ||| | ||| |||| ||||||||||
Db        960 ATCGGTCTACATCGCCACAGACTCTGAGAGCTACGTGTCAGAGATCCAGCAGCTCTTCAA 1019

Qy        903 AGGGAAGGTGAAGGTGGTGAGCCTGAAGCCTGAGGTGGCCCAGGTCGACCTGTACATCCT 962
              ||  ||||||| |||||||||||||| || |||||||||||| ||||||||||||||||
Db       1020 AGACAAGGTGAGGGTGGTGAGCCTGAAACCCGAGGTGGCCCAGATCGACCTGTACATCCT 1079

Qy        963 CGGCCAAGCCGACCACTTTATTGGCAACTGTGTCTCCTCCTTCACTGCCTTTGTGAAGCG 1022
              |||||| || |||||||| ||||| |||||||||||||| |||||||||| ||||||||
Db       1080 CGGCCAGGCTGACCACTTCATTGGAAACTGTGTCTCCTCGTTCACTGCCTTCGTGAAGCG 1139

Qy       1023 GGAGCGGGACCTCCAGGGGAGGCCGTCTTCTTTCTTCGGCATGGACAGGCCCCCTAAGCT 1082
              |||||||||||||| ||||||| ||| || ||||| |||||||||| ||| |  ||||
Db       1140 GGAGCGGGACCTCCATGGGAGGCAGTCGTCCTTCTTTGGCATGGACAGACCCTCCCAGCT 1199

Qy       1083 GCGGGACGAGTTCTGAT 1099
              ||||| || || ||||
Db       1200 TCGGGATGAATTTTGAT 1216


RESULT 5
BM909055
LOCUS       BM909055                1091 bp    mRNA    linear   EST 12-MAR-2002
DEFINITION  AGENCOURT_6611915 NIH_MGC_98 Homo sapiens cDNA clone IMAGE:5431669
            5', mRNA sequence.
ACCESSION   BM909055
VERSION     BM909055.1  GI:19359434
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1091)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
```

```
           Tissue Procurement: ATCC
            cDNA Library Preparation: Rubin Laboratory
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Agencourt Bioscience Corporation
            Clone distribution: MGC clone distribution information can be
           found through the I.M.A.G.E. Consortium/LLNL at:
           http://image.llnl.gov
           Plate: LLCM1904  row: e  column: 14
           High quality sequence stop: 629.
FEATURES             Location/Qualifiers
     source          1. .1091
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:5431669"
                     /tissue_type="astrocytoma grade IV, cell line"
                     /lab_host="DH10B (phage-resistant)"
                     /clone_lib="NIH_MGC_98"
                     /note="Organ: brain; Vector: pOTB7; Site_1: XhoI; Site_2:
                     EcoRI; cDNA made by oligo-dT priming.  Directionally
                     cloned into EcoRI/XhoI sites using the following 5'
                     adaptor: GGCACGAG(G).  Library constructed  by Ling Hong
                     in the laboratory of Gerald M. Rubin (University of
                     California, Berkeley) using ZAP-cDNA synthesis kit
                     (Stratagene) and Superscript II RT (Life Technologies).
                     Note: this is a NIH_MGC Library."
ORIGIN

  Query Match             67.6%;  Score 743.6;  DB 5;  Length 1091;
  Best Local Similarity   93.0%;  Pred. No. 1.3e-193;
  Matches  824;  Conservative    0;  Mismatches   54;  Indels    8;  Gaps    4;

Qy        178 CTCCATGTGTCCTACCAGAAGTACTTCAAGCTGGAGCCCCTCCAGGCTTACCATCGGGTC 237
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        132 CTCCATGTGTCCTACCAGAAGTACTTCAAGCTGGAGCCCCTCCAGGCTTACCATCGGGTC 191

Qy        238 ATCAGCTTGGAGGATTTCATGGAGAAGCTGGCACCCACCCACTGGCCCCCTGAGAAGCGG 297
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        192 ATCAGCTTGGAGGATTTCATGGAGAAGCTGGCACCCACCCACTGGCCCCCTGAGAAGCGG 251

Qy        298 GTGGCATACTGCTTTGAGGTGGCAGCCCAGCGAAGCCCAGATAAGAAGACGTGCCCCATG 357
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        252 GTGGCATACTGCTTTGAGGTGGCAGCCCAGCGAAGCCCAGATAAGAAGACGTGCCCCATG 311

Qy        358 AAGGAAGGAAACCCCTTTGGCCCATTCTGGGATCAGTTTCATGTGAGTTTCAACAAGTCG 417
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        312 AAGGAAGGAAACCCCTTTGGCCCATTCTGGGATCAGTTTCATGTGAGTTTCAACAAGTCG 371

Qy        418 GAGCTTTTTACAGGCATTTCCTTCAGTGCTTCCTACAGAGAACAATGGAGCCAGAGATTT 477
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        372 GAGCTTTTTACAGGCATTTCCTTCAGTGCTTCCTACAGAGAACAATGGAGCCAGAGATTT 431

Qy        478 TCTCCAAAGGAACATCCGGTGCTTGCCCTGCCAGGAGCCCCAGCCCAGTTCCCCGTCCTA 537
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        432 TCTCCAAAGGAACATCCGGTGCTTGCCCTGCCAGGAGCCCCAGCCCAGTTCCCCGTCCTA 491

Qy        538 GAAGAACACAGGCCACTACAGAAGTACATGGTATGGTCAGACGAAATGGTGAAGACGGGA 597
              || ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        492 GAGGAACACAGGCCACTACAGAAGTACATGGTATGGTCAGACGAAATGGTGAAGACGGGA 551

Qy        598 GAGGCCCAGATTCATGCCCACCTTGTCCGGCCCTATGTGGGCATTCATCTGCGCATTGGC 657
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        552 GAGGCCCAGATTCATGCCCACCTTGTCCGGCCCTATGTGGGCATTCATCTGCGCATTGGC 611

Qy        658 TCTGACTGGAAGAACGCCTGTGCCATGCTGAAGGACGGGACTGCAGGCTCGCACTTCATG 717
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        612 TCTGACTGGAAGAACGCCTGTGCCATGCTGAAGGACGGGACTGCAGGCTCGCACTTCATG 671

Qy        718 GCCTCTCCGCAGTGTGTGGGCTACAGCCGCAGCACAGCGGCCCCCCTCACGATGACTATG 777
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        672 GCCTCTCCGCAGTGTGTGGGCTACAGCCGCAGCACAGCGGCCCCCCTCACGATGACTATG 731

Qy        778 TGCCTGCCTGACCTGAAGGAGATCCAGAGGGCTGTGAAGCTCTGGGTGAGGTCGCTGGAT 837
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        732 TGCCTGCCTGACCTGAAGGAGATCCAGAGGGCTGTGAAGCTCTGGGTGAGGTCGCTGGAT 791

Qy        838 GCCCAGTCGGTCTACGTTGCTACTGATTCCGAGAGTTATGTGCCTGAGCTCCAACAGCTC 897
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        792 GCCCAGTCGGTCTACGTTGCTACTGATTCCGAGAGTTATGTGCCTGAAGCTCCACAGCTC 851

Qy        898 TTCAAA--GGGAAGGTGAAGGTGGTG--AGCCTGAAGCCTGAGGTGGCCCAGGTCGACCT 953
              || |||  || | || ||||||||||  | |   || |   ||| |||||||||| ||||
Db        852 TTTAAAAGGGAAGGGGGAAGGTGGTGGAACCTGAAACCTGTAGGGGGCCCAGGTCCACCT 911

Qy        954 GTACATCCTCGGCCAAGCCGACCACTTT-ATTGGCAACTGTGTCTCCTCCTTCACTG--- 1009
              ||| || ||  | |||||||| | |||| ||||| ||||||| | ||  | |  | |
Db        912 GTAAATTCTTCGGCAAGCCGAACCCTTTAATTGGGAACTGTGGCCCCCTCCTTCCCGGGC 971

Qy       1010 CCTTTGTGAAGCGGGAGCGGGACCTCCAGGGGAGGCCGTCTTCTTT 1055
              ||||||  ||  || | |||||||  |   || |   | |  ||||
Db        972 CCTTTGGAAAAGGGAAACGGGACCCTCTCAGGGGAGGGGCGGCTTT 1017


RESULT 6
BX345989
LOCUS       BX345989                841 bp    mRNA    linear   EST 23-APR-2004
DEFINITION  BX345989 Homo sapiens T CELLS (JURKAT CELL LINE) COT 10-NORMALIZED
            Homo sapiens cDNA clone CS0DJ008YF12 5-PRIME, mRNA sequence.
ACCESSION   BX345989
VERSION     BX345989.2  GI:46549776
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 841)
  AUTHORS   Li,W.B., Gruber,C., Jessee,J. and Polayes,D.
  TITLE     Full-length cDNA libraries and normalization
  JOURNAL   Unpublished (2001)
COMMENT     On May 5, 2003 this sequence version replaced gi:30377055.
            Contact: Genoscope
            Genoscope - Centre National de Sequencage
            2 rue Gaston Cremieux, CP 5706 - 91057 EVRY cedex - FRANCE
            Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr
            1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime
            end enriched, double-strand cDNA was digested with Not I and cloned
            into the Not I and EcoR V sites of the pCMVSPORT 6 vector. Library
            was normalized. Library was constructed by Life Technologies, a
            division of Invitrogen. This sequence belongs to sequence cluster
            9980.f
            For more information about this cluster, see
            http://www.genoscope.cns.fr/cdna?s=CS0AJ008DC06QP1&c=9980.f.
FEATURES             Location/Qualifiers
     source          1. .841
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="CS0DJ008YF12"
                     /cell_type="T CELLS (JURKAT CELL LINE) COT 10-NORMALIZED"
                     /cell_line="JURKAT"
                     /clone_lib="Homo sapiens T CELLS (JURKAT CELL LINE) COT
                     10-NORMALIZED"
                     /note="1st strand cDNA was primed with a NotI-oligo(dT)
                     primer. Five prime end enriched, double-strand cDNA was
                     digested with Not I and cloned into the Not I and EcoR V
                     sites of the pCMVSPORT 6 vector. Library was normalized."
ORIGIN

  Query Match             65.3%;  Score 718.4;  DB 5;  Length 841;
  Best Local Similarity   99.5%;  Pred. No. 1.1e-186;
  Matches  730;  Conservative    1;  Mismatches    2;  Indels    1;  Gaps    1;

Qy          1 ATGCCCGCGGGCTCCTGGGACCCGGCCGGTTACCTGCTCTACTGCCCCTGCATGGGGCGC 60
              ||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        109 ATGCCTGCGGGCTCCTGGGACCCGGCCGGTTACCTGCTCTACTGCCCCTGCATGGGGCGC 168
```

```
Qy         61 TTTGGGAACCAGGCCGATCACTTCTTGGGCTCTCTGGCATTTGCAAAGCTGCTAAACCGT 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        169 TTTGGGAACCAGGCCGATCACTTCTTGGGCTCTCTGGCATTTGCAAAGCTGCTAAACCGT 228

Qy        121 ACCTTGGCTGTCCCTCCTTGGATTGAGTACCAGCATCACAAGCCTCCTTTCACCAACCTC 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        229 ACCTTGGCTGTCCCTCCTTGGATTGAGTACCAGCATCACAAGCCTCCTTTCACCAACCTC 288

Qy        181 CATGTGTCCTACCAGAAGTACTTCAAGCTGGAGCCCCTCCAGGCTTACCATCGGGTCATC 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        289 CATGTGTCCTACCAGAAGTACTTCAAGCTGGAGCCCCTCCAGGCTTACCATCGGGTCATC 348

Qy        241 AGCTTGGAGGATTTCATGGAGAAGCTGGCACCCACCCACTGGCCCCCTGAGAAGCGGGTG 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        349 AGCTTGGAGGATTTCATGGAGAAGCTGGCACCCACCCACTGGCCCCCTGAGAAGCGGGTG 408

Qy        301 GCATACTGCTTTGAGGTGGCAGCCCAGCGAAGCCCAGATAAGAAGACGTGCCCCATGAAG 360
              |||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||||
Db        409 GCATACTGCTTTGAGGTGGCAGCCCAGCGAA-CCCAGATAAGAAGACGTGCCCCATGAAG 467

Qy        361 GAAGGAAACCCCTTTGGCCCATTCTGGGATCAGTTTCATGTGAGTTTCAACAAGTCGGAG 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        468 GAAGGAAACCCCTTTGGCCCATTCTGGGATCAGTTTCATGTGAGTTTCAACAAGTCGGAG 527

Qy        421 CTTTTTACAGGCATTTCCTTCAGTGCTTCCTACAGAGAACAATGGAGCCAGAGATTTTCT 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        528 CTTTTTACAGGCATTTCCTTCAGTGCTTCCTACAGAGAACAATGGAGCCAGAGATTTTCT 587

Qy        481 CCAAAGGAACATCCGGTGCTTGCCCTGCCAGGAGCCCCAGCCCAGTTCCCCGTCCTAGAA 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        588 CCAAAGGAACATCCGGTGCTTGCCCTGCCAGGAGCCCCAGCCCAGTTCCCCGTCCTAGAG 647

Qy        541 GAACACAGGCCACTACAGAAGTACATGGTATGGTCAGACGAAATGGTGAAGACGGGAGAG 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        648 GAACACAGGCCACTACAGAAGTACATGGTATGGTCAGACGAAATGGTGAAGACGGGAGAG 707

Qy        601 GCCCAGATTCATGCCCACCTTGTCCGGCCCTATGTGGGCATTCATCTGCGCATTGGCTCT 660
              |:||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        708 GSCCAGATTCATGCCCACCTTGTCCGGCCCTATGTGGGCATTCATCTGCGCATTGGCTCT 767

Qy        661 GACTGGAAGAACGCCTGTGCCATGCTGAAGGACGGGACTGCAGGCTCGCACTTCATGGCC 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        768 GACTGGAAGAACGCCTGTGCCATGCTGAAGGACGGGACTGCAGGCTCGCACTTCATGGCC 827

Qy        721 TCTCCGCAGTGTGT 734
              ||||||||||||||
Db        828 TCTCCGCAGTGTGT 841
```

RESULT 7
AL135434

```
LOCUS       AL135434                719 bp    mRNA    linear   EST 04-SEP-2003
DEFINITION  DKFZp762O0115_r1 762 (synonym: hmel2) Homo sapiens cDNA clone
            DKFZp762O0115 5', mRNA sequence.
ACCESSION   AL135434
VERSION     AL135434.1  GI:6603621
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 719)
  AUTHORS   Ansorge,W., Wirkner,U., Mewes,W., Weil,B. and Wiemann,S.
  TITLE     EST (Ansorge,W., Wirkner,U., Mewes,H.W., Weil,B. and Wiemann,S.)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: MIPS
            MIPS
            Ingolstaedter Landstr.1, D-85764 Neuherberg, Germany
            This is the 5' sequence of the clone insert
            Clone from S. Wiemann, Molecular Genome Analysis, German Cancer
            Research Center (DKFZ); Email s.wiemann@dkfz- heidelberg.de;
            sequenced by EMBL (European Molecular Biology Laboratories,
            Heidelberg/Germany) within the cDNA sequencing consortium of the
            German Genome Project.
            No s1 sequence available.
            This clone (DKFZp762O0115) is available at the RZPD in Berlin.
            Please contact the RZPD: Ressourcenzentrum, Heubnerweg 6, 14059
            Berlin-Charlottenburg, GERMANY; Email: clone@rzpd.de.
FEATURES             Location/Qualifiers
     source          1. .719
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="DKFZp762O0115"
                     /tissue_type="melanoma (MeWo cell line)"
                     /dev_stage="adult"
                     /lab_host="DH10B"
                     /clone_lib="762 (synonym: hmel2)"
                     /note="Vector: pSport1; Site_1: NotI; Site_2: SalI"
ORIGIN
```

```
  Query Match           65.2%;  Score 717.4;  DB 1;  Length 719;
  Best Local Similarity   99.9%;  Pred. No. 1.9e-186;
  Matches  718;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy        360 GGAAGGAAACCCCTTTGGCCCATTCTGGGATCAGTTTCATGTGAGTTTCAACAAGTCGGA 419
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 GGAAGGAAACCCCTTTGGCCCATTCTGGGATCAGTTTCATGTGAGTTTCAACAAGTCGGA 60

Qy        420 GCTTTTTACAGGCATTTCCTTCAGTGCTTCCTACAGAGAACAATGGAGCCAGAGATTTTC 479
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GCTTTTTACAGGCATTTCCTTCAGTGCTTCCTACAGAGAACAATGGAGCCAGAGATTTTC 120

Qy        480 TCCAAAGGAACATCCGGTGCTTGCCCTGCCAGGAGCCCCAGCCCAGTTCCCCGTCCTAGA 539
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 TCCAAAGGAACATCCGGTGCTTGCCCTGCCAGGAGCCCCAGCCCAGTTCCCCGTCCTAGA 180

Qy        540 AGAACACAGGCCACTACAGAAGTACATGGTATGGTCAGACGAAATGGTGAAGACGGGAGA 599
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 GGAACACAGGCCACTACAGAAGTACATGGTATGGTCAGACGAAATGGTGAAGACGGGAGA 240

Qy        600 GGCCCAGATTCATGCCCACCTTGTCCGGCCCTATGTGGGCATTCATCTGCGCATTGGCTC 659
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 GGCCCAGATTCATGCCCACCTTGTCCGGCCCTATGTGGGCATTCATCTGCGCATTGGCTC 300

Qy        660 TGACTGGAAGAACGCCTGTGCCATGCTGAAGGACGGGACTGCAGGCTCGCACTTCATGGC 719
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 TGACTGGAAGAACGCCTGTGCCATGCTGAAGGACGGGACTGCAGGCTCGCACTTCATGGC 360

Qy        720 CTCTCCGCAGTGTGTGGGCTACAGCCGCAGCACAGCGGCCCCCCCTCACGATGACTATGTG 779
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 CTCTCCGCAGTGTGTGGGCTACAGCCGCAGCACAGCGGCCCCCCCTCACGATGACTATGTG 420

Qy        780 CCTGCCTGACCTGAAGGAGATCCAGAGGGCTGTGAAGCTCTGGGTGAGGTCGCTGGATGC 839
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 CCTGCCTGACCTGAAGGAGATCCAGAGGGCTGTGAAGCTCTGGGTGAGGTCGCTGGATGC 480

Qy        840 CCAGTCGGTCTACGTTGCTACTGATTCCGAGAGTTATGTGCCTGAGCTCCAACAGCTCTT 899
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 CCAGTCGGTCTACGTTGCTACTGATTCCGAGAGTTATGTGCCTGAGCTCCAACAGCTCTT 540

Qy        900 CAAAGGGAAGGTGAAGGTGGTGAGCCTGAAGCCTGAGGTGGCCCAGGTCGACCTGTACAT 959
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 CAAAGGGAAGGTGAAGGTGGTGAGCCTGAAGCCTGAGGTGGCCCAGGTCGACCTGTACAT 600

Qy        960 CCTCGGCCAAGCCGACCACTTTATTGGCAACTGTGTCTCCTCCTTCACTGCCTTTGTGAA 1019
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 CCTCGGCCAAGCCGACCACTTTATTGGCAACTGTGTCTCCTCCTTCACTGCCTTTGTGAA 660

Qy       1020 GCGGGAGCGGGACCTCCAGGGGAGGCCGTCTTCTTTCTTCGGCATGGACAGGCCCCCTA 1078
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 GCGGGAGCGGGACCTCCAGGGGAGGCCGTCTTCTTTCTTCGGCATGGACAGGCCCCCTA 719
```

RESULT 8
BI524031

```
LOCUS       BI524031                889 bp    mRNA    linear   EST 29-AUG-2001
DEFINITION  603052202F1 NIH_MGC_122 Homo sapiens cDNA clone IMAGE:5201616 5',
            mRNA sequence.
ACCESSION   BI524031
VERSION     BI524031.1  GI:15348823
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 889)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Life Technologies, Inc.
             cDNA Library Preparation: Life Technologies, Inc.
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM11505  row: d  column: 01
            High quality sequence stop: 880.
FEATURES             Location/Qualifiers
     source          1. .889
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:5201616"
                     /lab_host="DH10B"
                     /clone_lib="NIH_MGC_122"
                     /note="Organ: pooled lung and spleen; Vector: pCMV-SPORT6;
                     Site_1: NotI; Site_2: EcoRV (destroyed); RNA source
                     anonymous pool of 24 week female lung, 16 week female
                     spleen, and 20-22 week male spleens.  Library is oligo-dT
                     primed and directionally cloned (EcoRV site is destroyed
                     upon  cloning).  Average insert size 1.4 kb, insert size
                     range 1-3 kb. Library is normalized and enriched for
                     full-length clones and was  constructed by C. Gruber
                     (Invitrogen).  Research Genetics tracking  code 026. Note:
                     this is a NIH_MGC Library."
ORIGIN

  Query Match             64.7%;  Score 712.2;  DB 4;  Length 889;
  Best Local Similarity   99.1%;  Pred. No. 5.6e-185;
  Matches  758;  Conservative    0;  Mismatches    3;  Indels    4;  Gaps    4;

Qy          1 ATGCCCGCGGGCTCCTGGGACCCGGCCGGTTACCTGCTCTACTGCCCCTGCATGGGGCGC 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         87 ATGCCCGCGGGCTCCTGGGACCCGGCCGGTTACCTGCTCTACTGCCCCTGCATGGGGCGC 146

Qy         61 TTTGGGAACCAGGCCGATCACTTCTTGGGCTCTCTGGCATTTGCAAAGCTGCTAAACCGT 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        147 TTTGGGAACCAGGCCGATCACTTCTTGGGCTCTCTGGCATTTGCAAAGCTGCTAAACCGT 206

Qy        121 ACCTTGGCTGTCCCTCCTTGGATTGAGTACCAGCATCACAAGCCTCCTTTCACCAACCTC 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        207 ACCTTGGCTGTCCCTCCTTGGATTGAGTACCAGCATCACAAGCCTCCTTTCACCAACCTC 266

Qy        181 CATGTGTCCTACCAGAAGTACTTCAAGCTGGAGCCCCTCCAGGCTTACCATCGGGTCATC 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        267 CATGTGTCCTACCAGAAGTACTTCAAGCTGGAGCCCCTCCAGGCTTACCATCGGGTCATC 326

Qy        241 AGCTTGGAGGATTTCATGGAGAAGCTGGCACCCACCCACTGGCCCCCTGAGAAGCGGGTG 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        327 AGCTTGGAGGATTTCATGGAGAAGCTGGCACCCACCCACTGGCCCCCTGAGAAGCGGGTG 386

Qy        301 GCATACTGCTTTGAGGTGGCAGCCCAGCGAAGCCCAGATAAGAAGACGTGCCCCATGAAG 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        387 GCATACTGCTTTGAGGTGGCAGCCCAGCGAAGCCCAGATAAGAAGACGTGCCCCATGAAG 446

Qy        361 GAAGGAAACCCCCTTTGGCCCATTCTGGGATCAGTTTCATGTGAGTTTCAACAAGTCGGAG 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        447 GAAGGAAACCCCCTTTGGCCCATTCTGGGATCAGTTTCATGTGAGTTTCAACAAGTCGGAG 506

Qy        421 CTTTTTACAGGCATTTCCTTCAGTGCTTCCTACAGAGAACAATGGAGCCAGAGATTTTCT 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        507 CTTTTTACAGGCATTTCCTTCAGTGCTTCCTACAGAGAACAATGGAGCCAGAGATTTTCT 566

Qy        481 CCAAAGGAACATCCGGTGCTTGCCCTGCCAGGAGCCCCAGCCCAGTTCCCCGTCCTAGAA 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        567 CCAAAGGAACATCCGGTGCTTGCCCTGCCAGGAGCCCCAGCCCAGTTCCCCGTCCTAGAG 626

Qy        541 GAACACAGGCCACTACAGAAGTACATGGTATGGTCAGACGAAATGGTGAAGACGGGAGAG 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        627 GAACACAGGCCACTACAGAAGTACATGGTATGGTCAGACGAAATGGTGAAGACGGGAGAG 686

Qy        601 GCCCAGATTCATGCCCACCTTGTCCGGCC-CTATGTGGGCATTCATCTGCGCATTGGCTC 659
              |||||||||||||||| || |||||||| ||||||||||||||||||||||||||||||
Db        687 GCCCAGATTCATGCCCAGCTGGTCCGGCCACTATGTGGGCATTCATCTGCGCATTGGCTC 746

Qy        660 TGACT-GGAAGAACGCCTGTGCCATGCTGAAGGACGGG-ACTGCAGGCTCGCACTTCATG 717
              ||||| |||||||||||||||||||||||||||||| ||||||||||||||||||||||
Db        747 TGACTGGGAAGAACGCCTGTGCCATGCTGAAGGACGGGAACTGCAGGCTCGCACTTCATG 806

Qy        718 GCCTCTCCGCAGTGTGTGGGCTACAGCCGCAGCACAGCGGCCCCC 762
              |||||||||||||||||||||||| |||||||||||||||||
Db        807 GCCTCTCCGCAGTGTGTGGGCTACAG-CGCAGCACAGCGGCCCCC 850


RESULT 9
BX449172
LOCUS       BX449172                898 bp    mRNA    linear   EST 06-MAY-2004
DEFINITION  BX449172 Homo sapiens FETAL LIVER Homo sapiens cDNA clone
            CS0DM014YK11 5-PRIME, mRNA sequence.
ACCESSION   BX449172
VERSION     BX449172.2  GI:47067286
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 898)
  AUTHORS   Li,W.B., Gruber,C., Jessee,J. and Polayes,D.
  TITLE     Full-length cDNA libraries and normalization
  JOURNAL   Unpublished (2001)
COMMENT     On May 22, 2003 this sequence version replaced gi:31030014.
            Contact: Genoscope
            Genoscope - Centre National de Sequencage
            2 rue Gaston Cremieux, CP 5706 - 91057 EVRY cedex - FRANCE
            Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr
            1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime
            end enriched, double-strand cDNA was digested with Not I and cloned
            into the Not I and EcoRV sites of the pCMVSPORT 6 vector. Library
            was not normalized. Library was constructed by Life Technologies, a
            division of Invitrogen.
            This sequence belongs to sequence cluster 9980.f
            For more information about this cluster, see
            http://www.genoscope.cns.fr/cdna?s=CS0AM014AF06QP1&c=9980.f.
FEATURES             Location/Qualifiers
     source          1. .898
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="CS0DM014YK11"
                     /tissue_type="FETAL LIVER"
                     /dev_stage="fetal"
                     /clone_lib="Homo sapiens FETAL LIVER"
                     /note="Organ: liver; Vector: pCMVSPORT_6; 1st strand cDNA
                     was primed with a NotI-oligo(dT) primer. Five prime end
                     enriched, double-strand cDNA was digested with Not I and
                     cloned into the Not I and EcoRV sites of the pCMVSPORT 6
                     vector. Library was not normalized."
ORIGIN
```

```
  Query Match             64.2%;  Score 706.6;  DB 5;  Length 898;
  Best Local Similarity   97.8%;  Pred. No. 2e-183;
  Matches  744;  Conservative    4;  Mismatches   10;  Indels    3;  Gaps    3;

Qy          1 ATGCCCGCGGGCTCCTGGGACCCGGCCGGTTACCTGCTCTACTGCCCCTGCATGGGGCGC 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        139 ATGCCCGCGGGCTCCTGGGACCCGGCCGGTTACCTGCTCTACTGCCCCTGCATGGGGCGC 198

Qy         61 TTTGGGAACCAGGCCGATCACTTCTTGGGCTCTCTGGCATTTGCAAAGCTGCTAAACCGT 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        199 TTTGGGAACCAGGCCGATCACTTCTTGGGCTCTCTGGCATTTGCAAAGCTGCTAAACCGT 258

Qy        121 ACCTTGGCTGTCCCTCCTTGGATTGAGTACCAGCATCACAAGCCTCCTTTCACCAACCTC 180
              ||||||||||||||||:||||||||||||||||||||||||||||||||||||||||||||
Db        259 ACCTTGGCTGTCCCTCHTTGGATTGAGTACCAGCATCACAAGCCTCCTTTCACCAACCTC 318

Qy        181 CATGTGTCCTACCAGAAGTACTTCAAGCTGGAGCCCCTCCAGGCTTACCATCGGGTCATC 240
              ||||||||||||||||||||||||:|||||||||||||||||||||||||||||||||||||
Db        319 CATGTGTCCTACCAGAAGTACTTYAAGCTGGAGCCCCTCCAGGCTTACCATCGGGTCATC 378

Qy        241 AGCTTGGAGGATTTCATGGAGAAGCTGGCACCCACCCACTGGCCCCCTGAGAAGCGGGTG 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        379 AGCTTGGAGGATTTCATGGAGAAGCTGGCACCCACCCACTGGCCCCCTGAGAAGCGGGTG 438

Qy        301 GCATACTGCTTTGAGGTGGCAGCCCAGCGAAGCCCAGATAAGAAGACGTGCCCCATGAAG 360
              |||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||
Db        439 GCATACTGCTTTGAGGTGGCAGCCCAGCGAA-CCCAGATAAGAAGACGTGCCCCATGAAG 497

Qy        361 GAAGGAAACCCCTTTGGCCCATTCTGGGATCAGTTTCATGTGAGTTTCAACAAGTCGGAG 420
              |||||||||||||||||||||||||||||||||||| ||| ||||||||||||||
Db        498 GAAGGAAACCCCTTTGGCCCATTCTGGGATCAGTTTCATGTAAGTGTCAACAAGTCGGAG 557

Qy        421 CTTTTTACAGGCATTTCCTTCAGTGCTTCCTACAGAGAACAATGGAGCCAGAGATTTTCT 480
              |||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||
Db        558 CTTTTTACAGGCATTTCCTTCAGTGCTTCGTACAGAGAACAATGGAGCCAGAGATTTTCT 617

Qy        481 CCAAAGGAACATCCGGTGCTTGCCCTGCCAGGAGCCCCAGCCCAGTTCCCCGTCCTAGAA 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        618 CCAAAGGAACATCCGGTGCTTGCCCTGCCAGGAGCCCCAGCCCAGTTCCCCGTCCTAGAG 677

Qy        541 GAACACAGGCCACTACAGAAGTACATGGTATGGTCAGACG-AAATGGTGAAGACGGGAGA 599
              |||||||||||||||||||||||||||||||||||||| ||||||||||||||||||||
Db        678 GAACACAGGCCACTACAGAAGTACATGGTATGGTCAGACGAAAATGGTGAAGACGGGAGA 737

Qy        600 GGCCCAGATTCATGCCCACCTTGTCCGGCCCTATGTGGGCATTCATCTGCGCATTGGCTC 659
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        738 CGCCCAGATTCATGCCCACCTTGTCCGGCCCTATGTGGGCATTCATCTGCGCATTGGCTC 797

Qy        660 TGACTGGAAGAACGCCTGTGCCATGCTGAAGGACGGGACTGCAGGCTCGCACTTCATGGC 719
              |||||||||||||||||| | ||||||||||||||||||||||||||||| ||||
Db        798 TGACTGGAAGAACGCCTGTCCAATGCTGAAGGACGGGACTGCAGGCTCGCACTTAATGGG 857

Qy        720 CT-CTCCGCAGTGTGTGGGCTACAGCCGCAGCACAGCGGCC 759
              || ||||||||||||||||:|||||||||||||| :||||
Db        858 CTYCTCCGCAGTGTGTGGGCDACAGCCGCAGCACACSGGCC 898


RESULT 10
CO774833
LOCUS       CO774833                942 bp    mRNA    linear   EST 04-AUG-2004
DEFINITION  ILLUMIGEN_MCQ_51950 Katze_MNLV Macaca nemestrina cDNA clone
            IBIUW:29960 5' similar to Bases 7 to 852 highly similar to human
            POFUT1 (Hs.178292), mRNA sequence.
ACCESSION   CO774833
VERSION     CO774833.1  GI:50976096
KEYWORDS    EST.
SOURCE      Macaca nemestrina (pig-tailed macaque)
  ORGANISM  Macaca nemestrina
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Cercopithecidae;
            Cercopithecinae; Macaca.
REFERENCE   1  (bases 1 to 942)
  AUTHORS   Katze,M.G., Thomas,M., Korth,M., Iadonato,S.P. and Magness,C.L.
  TITLE     Large-scale Rhesus Macaque cDNA Sequencing
  JOURNAL   Unpublished (2003)
COMMENT     Contact: C. Magness
            Illumigen Biosciences Inc.
            2203 Airport Way S, Suite 450, Seattle, WA 98134, USA
            Tel: 2063780400
            Fax: 2063780408
            Email: cmagness@illumigen.com
            Sequenced on 2004.07.21. 611 Q20 bases. Library Preparation: Prof.
            Michael Katze Lab at University of Washington DNA Sequencing:
            Illumigen Biosciences Inc. For further information, see
            http://www.macaque.org
            PCR PRimers
            FORWARD: CCCTCACTAAAGGGAACAAAA
            BACKWARD: CACTATAGGGCGAATTGGGTA
            Insert Length: 942   Std Error: 0.00
            Plate: CL000531  row: E  column: 09
            Seq primer: CCCTCACTAAAGGGAACAAAA
            POLYA=No.
FEATURES              Location/Qualifiers
     source           1. .942
                      /organism="Macaca nemestrina"
                      /mol_type="mRNA"
                      /strain="Indian"
                      /db_xref="taxon:9545"
                      /clone="IBIUW:29960"
                      /sex="male"
                      /lab_host="Electromax DH10B"
                      /clone_lib="Katze_MNLV"
                      /note="Organ: liver; Vector: pDONR 222; Site_1: BsrG I;
                      Site_2: BsrG I; Created from CloneMiner cDNA Library
                      Construction kit (catalog #18249-029)"
ORIGIN

  Query Match             61.2%;  Score 673.2;  DB 7;  Length 942;
  Best Local Similarity   94.1%;  Pred. No. 3.3e-174;
  Matches  743;  Conservative    0;  Mismatches   43;  Indels    4;  Gaps    4;

Qy          1 ATGCCCGCGGGCTCCTGGGACCCGGCCGGTTACCTGCTCTACTGCCCCTGCATGGGGCGC 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         95 ATGCCCGCGGGCTCCTGGGACCCGGCCGGTTACCTGCTCTACTGCCCCTGCATGGGGCGC 154

Qy         61 TTTGGGAACCAGGCCGATCACTTCTTGGGCTCTCTGGCATTTGCAAAGCTGCTAAACCGT 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        155 TTTGGGAACCAGGCCGATCACTTCTTGGGCTCTCTGGCATTTGCAAAGCTGCTAAACCGC 214

Qy        121 ACCTTGGCTGTCCCTCCTTGGATTGAGTACCAGCATCACAAGCCTCCTTTCACCAACCTC 180
              |||||||||||||||||||||||||||||||||||||||||||||||||| ||||||||||
Db        215 ACCTTGGCTGTCCCTCCTTGGATTGAGTACCAGCATCACAAGCCTCCCTTCACCAACCTC 274

Qy        181 CATGTGTCCTACCAGAAGTACTTCAAGCTGGAGCCCCTCCAGGCTTACCATCGGGTCATC 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        275 CATGTGTCCTACCAGAAGTACTTCAAGCTGGAGCCCCTCCAGGCTTACCATCGGGTCATC 334

Qy        241 AGCTTGGAGGATTTCATGGAGAAGCTGGCACCCACCCACTGGCCCCCTGAGAAGCGGGTG 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        335 AGCTTGGAGGATTTCATGGAGAAGCTGGCACCCACCCACTGGCCCCCTGAGAAGCGGGTG 394

Qy        301 GCATACTGCTTTGAGGTGGCAGCCCAGCGAAGCCCAGATAAGAAGACGTGCCCCATGAAG 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        395 GCATACTGCTTTGAGGTGGCAGCCCAGCGAAGCCCAGATAAGAAGACGTGCCCCATGAAG 454

Qy        361 GAAGGAAACCCCTTTGGCCCATTCTGGGATCAGTTTCATGTGAGTTTCAACAAGTCGGAG 420
              |||||||||||||||||||||||||||||||||||||||||||||||||||||| |||
Db        455 GAAGGAAACCCCTTTGGCCCATTCTGGGATCAGTTTCATGTGAGTTTCAACAAGTCTGAG 514

Qy        421 CTTTTTACAGGCATTTCCTTCAGTGCTTCCTACAGAGAACAATGGAGCCAGAGATTTTCT 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        515 CTTTTTACAGGCATTTCCTTCAGTGCTTCCTACAGAGAACAATGGAGCCAGAGATTTTCT 574

Qy        481 CCAAAGGAACATCCGGTGCTTGCCCTGCCAGGAGCCCCAGCCCAGTTCCCCGTCCTAGAA 540
              |||||||||||||||||||||||||||||| |||||||| |||||||||| ||||| |||
```

```
Db        575 CCAAAGGAACATCCGGTGCTTGCCCTGCCGGGAGCCCCGGCCCAGTTCCCTGTCCTGGAA 634

Qy        541 GAACACAGGCCACTACAGAAGTACATGGTATGGTCAGACGAAATGGTGAAGACGGGAGAG 600

Db        635 GAACACAGACCACTGCAAAAGTACATGGTATGGTCAGACCAAATGGTGAGGACGGGAGAA 694

Qy        601 GCCCAGATTCATGCCCACCTTGT-CCGGCCCTATGTGGGCATTCATCTGCGCATTGGC-T 658

Db        695 GCCCAGAATCCCGCCCACCTTGTCCCGGCCCTATGTGGGCATTCCTCTACGTAATGGCTT 754

Qy        659 CTGACTGGAAGAACGCCTGTGCCATGCTGAAGGACGGGACTGCAGGCTCGCACTTCATGG 718

Db        755 CTGACTGGAAAAATGCCTGTGCTATGCTAAAGGACGGAACTGGCAGCTCGACTTTCTTGG 814

Qy        719 CCTCTCCGCAGTGTGTGGG-CTACAGCCGCAGCACAG-CGGCCCCCCTCACGATGACTAT 776

Db        815 CCTCTCCCGAATGTGGGGGCCTACGCCCCCAACACAGCCGCCCCCCCTCCCGAAGACTAT 874

Qy        777 GTGCCTGCCT 786

Db        875 GTGCCTGCCT 884
```

```
RESULT 11
BG747760
LOCUS       BG747760                808 bp    mRNA    linear   EST 15-MAY-2001
DEFINITION  602705255F1 NIH_MGC_43 Homo sapiens cDNA clone IMAGE:4842061 5',
            mRNA sequence.
ACCESSION   BG747760
VERSION     BG747760.1  GI:14058413
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 808)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: ATCC
             cDNA Library Preparation: Ling Hong/Rubin Laboratory
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLCM1675  row: n  column: 14
            High quality sequence stop: 808.
FEATURES             Location/Qualifiers
     source          1. .808
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:4842061"
                     /tissue_type="normal pigmented retinal epithelium"
                     /lab_host="DH10B (phage-resistant)"
                     /clone_lib="NIH_MGC_43"
                     /note="Organ: eye; Vector: pOTB7; Site_1: XhoI; Site_2:
                     EcoRI; cDNA made by oligo-dT priming.  Directionally
                     cloned into EcoRI/XhoI sites using the following 5'
                     adaptor: GGCACGAG(G).  Library constructed  by Ling Hong
                     in the laboratory of Gerald M. Rubin (University of
                     California, Berkeley) using ZAP-cDNA synthesis kit
                     (Stratagene) and  Superscript II RT (Life Technologies).
                     Note: this is a NIH_MGC Library. |"
ORIGIN

  Query Match             60.2%;  Score 662.2;  DB 4;  Length 808;
  Best Local Similarity   95.3%;  Pred. No. 3.4e-171;
  Matches  693;  Conservative    0;  Mismatches   33;  Indels    1;  Gaps    1;
```

```
Qy          1 ATGCCCGCGGGGCTCCTGGGACCCGGCCGGTTACCTGCTCTACTGCCCCTGCATGGGGCGC 60

Db         82 ATGCCCGCGGGCTCCTGGGACCCGGCCGGTTACCTGCTCTACTGCCCCTGCATGGGGCGC 141

Qy         61 TTTGGGAACCAGGCCGATCACTTCTTGGGCTCTCTGGCATTTGCAAAGCTGCTAAACCGT 120

Db        142 TTTGGGAACCAGGCCGATCACTTCTTGGGCTCTCTGGCATTTGCAAAGCTGCTAAACCGT 201

Qy        121 ACCTTGGCTGTCCCTCCTTGGATTGAGTACCAGCATCACAAGCCTCCTTTCACCAACCTC 180

Db        202 ACCTTGGCTGTCCCTCCTTGGATTGAGTACCAGCATCACAAGCCTCCTTTCACCAACCTC 261

Qy        181 CATGTGTCCTACCAGAAGTACTTCAAGCTGGAGCCCCTCCAGGCTTACCATCGGGTCATC 240

Db        262 CATGTGTCCTACCAGAAGTACTTCAAGCTGGAGCCCCTCCAGGCTTACCATCGGGTCATC 321

Qy        241 AGCTTGGAGGATTTCATGGAGAAGCTGGCACCCACCCACTGGCCCCCTGAGAAGCGGGTG 300

Db        322 AGCTTGGAGGATTTCATGGAGAAGCTGGCACCCACCCACTGGCCCCCTGAGAAGCGGGTG 381

Qy        301 GCATACTGCTTTGAGGTGGCAGCCCAGCGAAGCCCAGATAAGAAGACGTGCCCCATGAAG 360

Db        382 GCATACTGCTTTGAGGTGGCAGCCCAGCGAAGCCCAGATAAGAAGACGTGCCCCATGAAG 441

Qy        361 GAAGGAAACCCCTTTGGCCCATTCTGGGATCAGTTTCATGTGAGTTTCAACAAGTCGGAG 420

Db        442 GAAGGAAACCCCTTTGGCCCATTCTGGGATCAGTTTCATGTGAGTTTCAACAAGTCGGAG 501

Qy        421 CTTTTTACAGGCATTTCCTTCAGTGCTTCCTACAGAGAACAATGGAGCCAGAGATTTTCT 480

Db        502 CTTTTTACAGGCATTTCCTTCAGTGCTTCCTACAGAGAACAATGGAGCCAGAGATTTTCT 561

Qy        481 CCAAAGGAACATCCGGTGCTTGCCCTGCCAGGAGCCCCAGCCCAGTTCCCCGTCCTAGAA 540

Db        562 CCAAAGGAACATCCGGTGCTTGCCCTGCCAGGAGCCCCAGCCCAGTTCCCCGTCCTAGAG 621

Qy        541 GAACACAGGCCACTACAGAAGTACATGGTATGGTCAGACGAAATGGTGAAGAC-GGGAGA 599

Db        622 GAACACAGGCCACTACAGAAGTACATGGTATGGTCAGACGAAATGGTGAAGACGGGGAGA 681

Qy        600 GGCCCAGATTCATGCCCACCTTGTCCGGCCCTATGTGGGCATTCATCTGCGCATTGGCTC 659

Db        682 GGCCCAGATTCATGCCCAACTGGTCCGGACCTATGTTGGGCATCAATCTGCGCATTGGTC 741

Qy        660 TGACTGGAAGAACGCCTGTGCCATGCTGAAGGACGGGACTGCAGGCTCGCACTTCATGGC 719

Db        742 TGACTGGAAGAACGCCTGTGCCATGCTGAAGGACGGGGACTGAGGCTCGGCACTTCATGG 801

Qy        720 CTCTCCG 726

Db        802 CTCTCCG 808
```

```
RESULT 12
BP144264
LOCUS       BP144264                801 bp    mRNA    linear   EST 30-DEC-2003
DEFINITION  BP144264 full-length enriched swine cDNA library, adult ovary Sus
            scrofa cDNA clone OVR010034G09 5', mRNA sequence.
ACCESSION   BP144264
VERSION     BP144264.1  GI:40393735
KEYWORDS    EST.
SOURCE      Sus scrofa (pig)
  ORGANISM  Sus scrofa
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
REFERENCE   1  (bases 1 to 801)
  AUTHORS   Uenishi,H., Eguchi,T., Suzuki,K., Sawazaki,T., Toki,D., Shinkai,H.,
            Okumura,N., Hamasima,N. and Awata,T.
  TITLE     PEDE (Pig EST Data Explorer): construction of a database for ESTs
            derived from porcine full-length cDNA libraries
  JOURNAL   Nucleic Acids Res. 32 (1), D484-D488 (2004)
COMMENT     Contact: Hirohide Uenishi
            Animal Genome Laboratory, Genome Research Department
```

```
           National Institute of Agrobiological Sciences
           2 Ikenodai, Tsukuba, Ibaraki 305-8602, Japan
           Tel: +81-29-838-8627
           Fax: +81-29-838-8627
           Email: huenishi@affrc.go.jp
           EST project with full-length enriched cDNA libraries carried out in
           Animal Genome Research Program (Japan) by National Institute of
           Agrobiological Sciences and STAFF-Institute
           Single pass sequencing of clones derived from oligo-capped cDNA
           library
           Vector sequences were eliminated by RepeatMasker version 2002/07/13
           and crossmatch version 0.990319
           Low quality bases were trimmed based on the quality values.
FEATURES             Location/Qualifiers
     source          1. .801
                     /organism="Sus scrofa"
                     /mol_type="mRNA"
                     /db_xref="taxon:9823"
                     /clone="OVR010034G09"
                     /tissue_type="ovary"
                     /dev_stage="adult"
                     /clone_lib="full-length enriched swine cDNA library, adult
                     ovary"
ORIGIN

  Query Match             60.0%;  Score 660.2;  DB 5;  Length 801;
  Best Local Similarity   89.0%;  Pred. No. 1.2e-170;
  Matches  713;  Conservative    0;  Mismatches   88;  Indels    0;  Gaps    0;

Qy        194 AGAAGTACTTCAAGCTGGAGCCCCTCCAGGCTTACCATCGGGTCATCAGCTTGGAGGATT 253
              | ||||||||||||||||||||||||||||  ||||| |||||||||||| ||||||| |
Db          1 AAAAGTACTTCAAGCTGGAGCCCCTCCAGGTCTACCACCGGGTCATCAGCCTGGAGGACT 60

Qy        254 TCATGGAGAAGCTGGCACCCACCCACTGGCCCCCTGAGAAGCGGGTGGCATACTGCTTTG 313
              |||||||||||||||||||||||||||||||||||||||| |||||||||||||| ||||
Db         61 TCATGGAGAAGCTGGCACCCACCCACTGGCCCCCTGAGAGGCGGGTGGCATACTGTTTTG 120

Qy        314 AGGTGGCAGCCCAGCGAAGCCCAGATAAGAAGACGTGCCCCATGAAGGAAGGAAACCCCT 373
              ||||||||| |||||||||||||||||||||| ||||||||||||||||||||||||||
Db        121 AGGTGGCAGCACAGCGAAGCCCAGATAAGAAGACATGCCCCATGAAGGAAGGAAACCCCT 180

Qy        374 TTGGCCCATTCTGGGATCAGTTTCATGTGAGTTTCAACAAGTCGGAGCTTTTTACAGGCA 433
              |||||||||| |||||||||||||||||||||||||||||| || |||||| ||||||
Db        181 TTGGCCCATTTTGGGATCAGTTTCATGTGAGTTTCAACAAGTCAGAACTTTTTGCAGGCA 240

Qy        434 TTTCCTTCAGTGCTTCCTACAGAGAACAATGGAGCCAGAGATTTTCTCCAAAGGAACATC 493
              ||||||||||||| ||||||| ||| || |||| ||| ||||||||||||||||||||||
Db        241 TTTCCTTCAGTGCCTCCTACAAAGACCAGTGGATCCAAAGATTTTCTCCAAAGGAACATC 300

Qy        494 CGGTGCTTGCCCTGCCAGGAGCCCCAGCCCAGTTCCCCGTCCTAGAAGAACACAGGCCAC 553
              | |||||||||||||||| ||||| || ||||||||||||| || ||||| |||||||
Db        301 CAGTGCTTGCCCTGCCAGGGGCCCCGGCTCAGTTCCCCGTCCTGGAGGAACATAGGCCAC 360

Qy        554 TACAGAAGTACATGGTATGGTCAGACGAAATGGTGAAGACGGGAGAGGCCCAGATTCATG 613
              | ||||| |||||||| |||||||||| ||||||| ||| || |||||||||| | ||
Db        361 TTCAGAAATACATGGTGTGGTCAGACGAGATGGTGAGGACAGGGGAGGCCCAGATCCGTG 420

Qy        614 CCCACCTTGTCCGGCCCTATGTGGGCATTCATCTGCGCATTGGCTCTGACTGGAAGAACG 673
              ||||||| ||||||||| |||||||||| ||||||||||||||||||||||||||||||
Db        421 CCCACCTCATCCGGCCCTACGTGGGCATTCACCTGCGCATTGGCTCTGACTGGAAGAACG 480

Qy        674 CCTGTGCCATGCTGAAGGACGGGACTGCAGGCTCGCACTTCATGGCCTCTCCGCAGTGTG 733
              | || |||||||||||||||| ||||| ||| | |||||||||||||| ||||||||  |
Db        481 CGTGCGCCATGCTGAAGGACGGCACTGCCGGCGCTCACTTCATGGCCTCCCCGCAGTGCG 540

Qy        734 TGGGCTACAGCCGCAGCACAGCGGCCCCCCTCACGATGACTATGTGCCTGCCTGACCTGA 793
              |||| |||||||||  ||| || ||||| ||||| ||||| |||||||| || |||||||
Db        541 TGGGTTACAGCCGCCACACCGCCGCCCCGCTCACCATGACCATGTGCCTCCCCGACCTGA 600

Qy        794 AGGAGATCCAGAGGGCTGTGAAGCTCTGGGTGAGGTCGCTGGATGCCCAGTCGGTCTACG 853
              ||||||||  | | ||  |||||||||||||| | |||||   |||||||| ||||||
Db        601 AGGAGATCAGGCGCGCCCTGAAGCTCTGGGTGACGGCGCTGAGCGCCCAGTCCGTCTACA 660

Qy        854 TTGCTACTGATTCCGAGAGTTATGTGCCTGAGCTCCAACAGCTCTTCAAAGGGAAGGTGA 913
              | || || || || |||||||||  |||| ||| |||| ||||||||||||| |||||||
Db        661 TCGCCACGGACTCTGAGAGTTACCTGCCGGAGATCCAGCAGCTCTTCAAAGGAAAGGTGA 720

Qy        914 AGGTGGTGAGCCTGAAGCCTGAGGTGGCCCAGGTCGACCTGTACATCCTCGGCCAAGCCG 973
              |||||||||| ||||| |||||||||||||| |||||||||||||| ||||| || |
Db        721 AGGTGGTGAGTCTGAAACCTGAGGTGGCCCAGATCGACCTGTACATCCTTGGCCAGGCTG 780

Qy        974 ACCACTTTATTGGCAACTGTG 994
              |||||||||||||||||||||
Db        781 ACCACTTTATTGGCAACTGTG 801


RESULT 13
BQ964179
LOCUS       BQ964179                988 bp    mRNA    linear   EST 21-AUG-2002
DEFINITION  AGENCOURT_10055799 NIH_MGC_134 Mus musculus cDNA clone
            IMAGE:6509783 5', mRNA sequence.
ACCESSION   BQ964179
VERSION     BQ964179.1  GI:22379657
KEYWORDS    EST.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 988)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Dr. David Rowe
             cDNA Library Preparation: Invitrogen Corp
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Agencourt Bioscience Corporation
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM14077  row: j  column: 24
            High quality sequence stop: 702.
FEATURES             Location/Qualifiers
     source          1. .988
                     /organism="Mus musculus"
                     /mol_type="mRNA"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:6509783"
                     /tissue_type="undifferentiated limb"
                     /lab_host="DH10B (phage-resistant)"
                     /clone_lib="NIH_MGC_134"
                     /note="Vector: pCMV-SPORT6.1; Site_1: EcoRV; Site_2: NotI;
                     Cloned unidirectionally. Primer: Oligo dT. Average insert
                     size 1.7 kb. Constructed by ResGen, Invitrogen Corp. Note:
                     this is a NIH_MGC Library."
ORIGIN

  Query Match             58.1%;  Score 639;  DB 5;  Length 988;
  Best Local Similarity   87.3%;  Pred. No. 9e-165;
  Matches  721;  Conservative    0;  Mismatches  103;  Indels    2;  Gaps    2;

Qy          3 GCCCGCGGGCTCCTGGGACCCGGCCGGTTACCTGCTCTACTGCCCCTGCATGGGGCGCTT 62
              | | |||||||||||||||| |||||||||||||||||||| ||||||||||||||||||
Db        148 GTCAGCGGGCTCCTGGGACCTGGCCGGTTACCTGCTCTACTGTCCCTGCATGGGGCGCTT 207

Qy         63 TGGGAACCAGGCCGATCACTTCTTGGGCTCTCTGGCATTTGCAAAGCTGCTAAACCGTAC 122
              |||||||||||| |||||||||||||||| |||||||||| |||||||| ||||| ||
Db        208 TGGGAACCAGGCTGATCACTTCTTGGGCTCCCTGGCATTTGCGAAGCTGCTGAACCGCAC 267

Qy        123 CTTGGCTGTCCCTCCTTGGATTGAGTACCAGCATCACAAGCCTCCTTTCACCAACCTCCA 182
              |||||||| ||||| |||||||| ||||| ||||||||||||||||||||||||||||||
Db        268 CTTGGCTGTACCTCCATGGATTGAATACCAACATCACAAGCCTCCTTTCACCAACCTCCA 327

Qy        183 TGTGTCCTACCAGAAGTACTTCAAGCTGGAGCCCCTCCAGGCTTACCATCGGGTCATCAG 242
              |||||||||||| |||||||||| |||||||| ||||| || |||||||||||  ||||
```

```
Db        328 TGTGTCCTACCAAAAGTACTTCAAACTGGAGCCTCTCCAAGCCTACCATCGGGTTGTCAG 387

Qy        243 CTTGGAGGATTTCATGGAGAAGCTGGCACCCACCCACTGGCCCCCTGAGAAGCGGGTGGC 302

Db        388 CCTGGAGGACTTCATGGAAAATCTGGCACCCTCCCACTGGCCCCCTGAGAAGCGAGTGGC 447

Qy        303 ATACTGCTTTGAGGTGGCAGCCCAGCGAAGCCCAGATAAGAAGACGTGCCCCATGAAGGA 362

Db        448 ATACTGCTTTGAGGTGGCAGCCCAGCGAAGTCCTGATAAGAAGACATGTCCCATGAAGGA 507

Qy        363 AGGAAACCCCTTTGGCCCATTCTGGGATCAGTTTCATGTGAGTTTCAACAAGTCGGAGCT 422

Db        508 AGGAAATCCTTTTGGGCCATTCTGGGACCAGTTTCATGTGAGTTTCAATAAGTCAGAACT 567

Qy        423 TTTTACAGGCATTTCCTTCAGTGCTTCCTACAGAGAACAATGGAGCCAGAGATTTTCTCC 482

Db        568 GTTCACAGGCATTTCCTTCAGCGCCTCCTACAAAGAACAATGGACCCAGAGATTTCCTGC 627

Qy        483 AAAGGAACATCCGGTGCTTGCCCTGCCAGGAGCCCCAGCCCAGTTCCCCGTCCTAGAAGA 542

Db        628 AAAAGAGCATCCTGTGCTCGCACTGCCTGNGGCCCCAGCACAGTTCCCTGTCCTGGAGGA 687

Qy        543 ACACAGGCCACTACAGAAGTACATGGTATGGTCAGACGAAATGGTGAAGACGGGAGAGGC 602

Db        688 ACACAGGGAGCTCCAGAAGTACATGGTGTGGTCAGATGAGATGGTGAGGACGGGAGAGGC 747

Qy        603 CCAGATTCATGCCCACCTTGTCCGGCCCTATGTGGGCATTCATCTGCGCATTGGCTCTGA 662

Db        748 CCTGATCAGTGCCCACCTCGTCCGGCCCTATGTGGGCATTCATCTGCGCATTGGCTCCGA 807

Qy        663 CTGGAAGAACGCCTGTGCCATGCTGAAGGACGGGACTGCAGGCTCGCACTTCATGGCCTC 722

Db        808 CTGGAAGAATGCCTGTGCCATGCTGGAGGATGGAACTGCAGGGTCACACTTCATGGCTTC 867

Qy        723 TCC-GCAGTGTGTGGGCTACAGCCGCAGCACAGCGGCCCCCCTCACGATGACTATGTGCC 781

Db        868 CCCTTCNNTGTGTGGGCTATAGCCGCAGCACAGCGACCCCTCTCAC-ATGACCATGTGCC 926

Qy        782 TGCCTGACCTGAAGGAGATCCAGAGGGCTGTGAAGCTCTGGGTGAG 827

Db        927 TCCCTGACCTGAAGGAAATCAGCGGGCTGGGGAACGCTTTGGGGTG 972


RESULT 14
BI757540
LOCUS       BI757540                874 bp    mRNA    linear   EST 25-SEP-2001
DEFINITION  603029693F1 NIH_MGC_114 Homo sapiens cDNA clone IMAGE:5199862 5',
            mRNA sequence.
ACCESSION   BI757540
VERSION     BI757540.1  GI:15749118
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 874)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Life Technologies, Inc.
             cDNA Library Preparation: Life Technologies, Inc.
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM11500  row: j  column: 23
            High quality sequence stop: 798.
FEATURES             Location/Qualifiers
     source          1. .874
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:5199862"
                     /lab_host="DH10B"
                     /clone_lib="NIH_MGC_114"
                     /note="Organ: brain; Vector: pCMV-SPORT6; Site_1: NotI;
                     Site_2: EcoRV (destroyed); RNA source anonymous pool of 6
                     male brains, age range 23-27 yo. Library is oligo-dT
                     primed and directionally cloned (EcoRV site is destroyed
                     upon cloning). Average insert size 1.5 kb, insert size
                     range 1-3 kb. Library is normalized and enriched for
                     full-length clones and was constructed by C. Gruber
                     (Invitrogen). Research Genetics tracking code 019. Note:
                     this is a NIH_MGC Library."
ORIGIN

  Query Match             57.9%;  Score 636.4;  DB 4;  Length 874;
  Best Local Similarity   96.1%;  Pred. No. 4.5e-164;
  Matches  717;  Conservative    0;  Mismatches   21;  Indels    8;  Gaps    6;

Qy          1 ATGCCCGCGGGCTCCTGGGACCCGGCCGGTTACCTGCTCTACTGCCCCTGCATGGGGCGC 60

Db        109 ATGCCCGCGGGCTCCTGGGACCCGGCCGGTTACCTGCTCTACTGCCCCTGCATGGGGCGC 168

Qy         61 TTTGGGAACCAGGCCGATCACTTCTTGGGCTCTCTGGCATTTGCAAAGCTGCTAAACC-G 119

Db        169 TTTGGGAACCAGGCCGATCACTTCTTGGGCTCTCTGGCATTTGCAAAGCTGCTAAACCTG 228

Qy        120 TACCTTGGCTGTCCCTCCTTGGATTGAGTACCAGCATCACAAGCCTCCTTTCACCAACCT 179

Db        229 TACCTTGGCTGTCCCTCCTTGGATTGAGTACCAGCATCACAAGCCTCCTTTCACCAACCT 288

Qy        180 CCATGTGTCCTACCAGAAGTACTTCAAGCTGGAGCCCCTCCAGGCTTACCATCGGGTCAT 239

Db        289 CCATGTGTCCTACCAGAAGTACTTCAAGCTGGAGCCCCTCCAGGCTTACCATCGGGTCAT 348

Qy        240 CAGCTTGGAGGATTTCATGGAGAAGCTGGCACCCACCCACTGGCCCCCTGAGAAGCGGGT 299

Db        349 CAGCTTGGAGGATTTCATGGAGAAGCTGGCACCCACCCACTGGCCCCCTGAGAAGCGGGT 408

Qy        300 GGCATACTGCTTTGAGGTGGCAGCCCAGCGAAGCCCAGATAAGAAGACGTGCCCCATGAA 359

Db        409 GGCATACTGCTTTGAGGTGGCAGCCCAGCGAAGCCCAGATAAGAAGACGTGCCCCATGAA 468

Qy        360 GGAAGGAAACCCCTTTGGCCCATTCTGGGATCAGTTTCATGTGAGTTTCAACAAGTCGGA 419

Db        469 GGAAGGAAACCCCTTTGGCCCATTCTGGGATCAG-TTCATGTGAGTTTCAACAAGTCGGA 527

Qy        420 GCTTTTTACAGGCATTTCCTTCAGTGCTTCCTACAGAGAACAATGGAGCCAGAGATTTTC 479

Db        528 GCTTTTTACAGGCATTTCCTTCAGTGCTTCCTACAGAGAACAATGGAGCCAGAGATTTTC 587

Qy        480 TCCAAAGGAACATCCGGTGCTTGCCCTGCCAGGAGCCCC-AGCCCAGTTCCCCGTCCTAG 538

Db        588 TCCAAAGGAACATCCGGTGCTTGCCCTGCCAGGAGCCCCAAGCCCAGTTCCCCGTCCTAG 647

Qy        539 AAGAACACAGGCCACTACAGAAGTACATGGTATGGTCAGAC--GAAATGGTGAAGACGGG 596

Db        648 AGGAACACAGGCCACTACAGAAGTACATGGTATGGTCAGACCGACAATGGTGAAGACGGG 707

Qy        597 AGAGGCCCAGATTCATGCCCACCTTGTCCGGCCCTATGTGGGCATTCATCTGCGCATTGG 656

Db        708 AGAGGCCCAGATTCATGCCCACCTTGTCCGGCCCTATGTGGGCATTCATCTGCGCATTGC 767

Qy        657 CT-CTGACTGGAAGAACGCCTGTGCCATGCTGA--AGGACGGGACTGCAGGCTCGCACTT 713

Db        768 CTCCTGACTGGAGGAACGCCTGTGCCATGCTGACCGGACTGGGCCTGCGGGCTCCCCTTC 827

Qy        714 CATGGCCTCTCCGCAGTGTGTGGGCT 739

Db        828 ATGGCCTCCTCCGCAGTGTGTGGGCT 853
```

```
RESULT 15
BI410783
LOCUS       BI410783                947 bp    mRNA    linear   EST 14-AUG-2001
DEFINITION  602963556F1 NCI_CGAP_Lu33 Mus musculus cDNA clone IMAGE:5119332 5',
            mRNA sequence.
ACCESSION   BI410783
VERSION     BI410783.1  GI:15171706
KEYWORDS    EST.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 947)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Gilbert Smith, Ph.D.
             cDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima
            Bonaldo, Ph.D.
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by:Incyte Genomics, Inc.
             Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM11290  row: o  column: 13
            High quality sequence start: 21
            High quality sequence stop: 919.
FEATURES             Location/Qualifiers
     source          1. .947
                     /organism="Mus musculus"
                     /mol_type="mRNA"
                     /strain="CZECH II"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:5119332"
                     /tissue_type="pooled lung tumors"
                     /lab_host="DH10B (phage-resistant)"
                     /clone_lib="NCI_CGAP_Lu33"
                     /note="Organ: lung; Vector: pT7T3D-Pac (Pharmacia) with a
                     modified polylinker; Site_1: NotI; Site_2: EcoRI; 1st
                     strand cDNA was prepared from mRNA obtained from pooled
                     lung tumors with a Not I - oligo(dT) primer [5'
                     TGTTACCAATCTGAAGTGGGAGCGGCCGCCTCTGTTTTTTTTTTTTTTTT 3'].
                     Double-stranded cDNA was ligated to Eco RI adaptors
                     (Pharmacia), digested with Not I and cloned into the Not
                     I and Eco RI sites of the modified pT7T3 vector. Library
                     went through one round of normalization, and was
                     constructed by Bento Soares and M. Fatima Bonaldo. "
ORIGIN
```

Query Match 57.3%; Score 630.8; DB 4; Length 947;
Best Local Similarity 86.9%; Pred. No. 1.6e-162;
Matches 763; Conservative 0; Mismatches 107; Indels 8; Gaps 6;

```
Qy          3 GCCCGCGGGCTCCTGGGACCCGGCCGGTTACCTGCTCTACTGCCCCTGCATGGGGCGCTT 62
              | | |||||||||||||||| |||||||||||||||||||||| ||||||||||||||||||
Db         70 GTCAGCGGGCTCCTGGGACCTGGCCGGTTACCTGCTCTACTGTCCCTGCATGGGGCGCTT 129

Qy         63 TGGGAACCAGGCCGATCACTTCTTGGGCTCTCTGGCATTTGCAAAGCTGCTAAACCGTAC 122
              |||||||||||| |||||||||||||||| |||||||||| |||||||| ||||| ||
Db        130 TGGGAACCAGGCTGATCACTTCTTGGGCTCCCTGGCATTTGCGAAGCTGCTGAACCGCAC 189

Qy        123 CTTGGCTGTCCCTCCTTGGATTGAGTACCAGCATCACAAGCCTCCTTTCACCAACCTCCA 182
              ||||||||| ||||| |||||||| ||||| ||||||||||||||||||||||||||||
Db        190 CTTGGCTGTACCTCCATGGATTGAATACCAACATCACAAGCCTCCTTTCACCAACCTCCA 249

Qy        183 TGTGTCCTACCAGAAGTACTTCAAGCTGGAGCCCCTCCAGGCTTACCATCGGGTCATCAG 242
              |||||||||||| |||||||||| |||||||| ||||| || ||||||||||   ||||
Db        250 TGTGTCCTACCAAAAGTACTTCAAACTGGAGCCTCTCCAAGCCTACCATCGGGTTGTCAG 309

Qy        243 CTTGGAGGATTTCATGGAGAAGCTGGCACCCACCCACTGGCCCCCTGAGAAGCGGGTGGC 302
              | ||||||| |||||||| || ||||||||| |||||||||||||||||||||||| |||||

Db        310 CCTGGAGGACTTCATGGAAAATCTGGCACCCTCCCACTGGCCCCCTGAGAAGCGAGTGGC 369

Qy        303 ATACTGCTTTGAGGTGGCAGCCCAGCGAAGCCCAGATAAGAAGACGTGCCCCATGAAGGA 362
              |||||||||||||||||||||||||||||| || |||||||||| || ||||||||||
Db        370 ATACTGCTTTGAGGTGGCAGCCCAGCGAAGTCCTGATAAGAAGACATGTCCCATGAAGGA 429

Qy        363 AGGAAACCCCTTTGGCCCATTCTGGGATCAGTTTCATGTGAGTTTCAACAAGTCGGAGCT 422
              |||||| || ||||| |||||||||| |||||||||||||||||||| ||||| || ||
Db        430 AGGAAATCCTTTTGGGCCATTCTGGGACCAGTTTCATGTGAGTTTCAATAAGTCAGAACT 489

Qy        423 TTTTACAGGCATTTCCTTCAGTGCTTCCTACAGAGAACAATGGAGCCAGAGATTTTCTCC 482
              || |||||||||||||||| || ||||||| |||||||||| ||||||||| || |
Db        490 GTTCACAGGCATTTCCTTCAGCGCCTCCTACAAAGAACAATGGACCCAGAGATTTCCTGC 549

Qy        483 AAAGGAACATCCGGTGCTTGCCCTGCCAGGAGCCCCAGCCCAGTTCCCCGT-CCTAGAAG 541
              ||| || ||||| ||||| || ||||| || |||||||| |||||||| || ||| || |
Db        550 AAAAGAGCATCCTGTGCTCGCACTGCCTGGGGCCCCAGCACAGTTCCCTGTCCCTGGAGG 609

Qy        542 AACACAGGCCACTACAGAAGTACATGGTATGGTCAGACGAAATGGTGAAGACGGGAGAGG 601
              ||||||||   || |||||||||||||| |||||||| || ||||||| ||||||||||
Db        610 AACACAGGGAGCTCCAGAAGTACATGGTGTGGTCAGATGAGATGGTGAGGACGGGAGAGG 669

Qy        602 CCCAGATTCATGCCCCACCTTGTCCGGCCCTATGTGGGCATTCATCTGCGCATTGGCTCTG 661
              ||| |||   |||||||| |||||||||||||||||||||||||||| |||||| |
Db        670 CCCTGATCAGTGCCCACCTCGTCCGGCCCTATGTGGGCATTCATCTGCGCA-TGGCTCCG 728

Qy        662 ACTGGAAGAACGCCTGTGCCATGCTGAAGGACGGGACTGCAGGCTCGCACTTCATGGCCT 721
              |||||||||| |||||||||||||||||| || |||||||||| |||||||||| |
Db        729 ACTGGAAGAATGCCTGTGCCATGCTGAAGGATGGAACTGCAGGCTCACACTTCATGGCTT 788

Qy        722 CTCCGCAGTGTGTGGGCTACAGCCGCAGCACAGCGGCCCCCCTCACGATGACTATGTGCC 781
              | || |||||||||||||| |||||||||| | ||| |       ||||| |||||||
Db        789 CCCCTCAGTGTGTGGGCTATAGCCGCAGCACAAGGACCCTC--TCACATGACCATGTGCC 846

Qy        782 TGCCTGACCTGAAGGAGATCCAGAGGGCTGTGAAGCTCTGGGTGAGGTCGCTGGATGCCC 841
              | |||||||||||||| || ||| ||||||||| ||| ||||||||  |  ||    ||
Db        847 TCCCTGACCTGAAGGAAAT-CAGCGGGCTGTGACGCTTTGGGTGAGAGCATTG--ATGCC 903

Qy        842 AGTCGGTCTACGTTGCTACTGATTCCGAGAGTTATGTG 879
              | ||||||||| | || || || || ||||| || |||
Db        904 AATCGGTCTACATCGC-ACAGACTCTGAGAGCTACGTG 940
```

Search completed: November 9, 2005, 10:38:07
Job time : 3990.87 secs

GenCore version 5.1.6

OM nucleic - nucleic search, using sw model

Run on: November 9, 2005, 08:00:04 ; Search time 984.698 Seconds
(without alignments)
9238.165 Million cell updates/sec

Title: US-09-774-954-16
Perfect score: 1100
Sequence: 1 ATGCCCGCGGGCTCCTGGGA.....CTGCGGGACGAGTTCTGATT 1100

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 9794790 seqs, 4134909567 residues

Total number of hits satisfying chosen parameters: 19589580

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published_Applications_NA:*
1: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq:*
2: /cgn2_6/ptodata/1/pubpna/PCT_NEW_PUB.seq:*
3: /cgn2_6/ptodata/1/pubpna/US06_NEW_PUB.seq:*
4: /cgn2_6/ptodata/1/pubpna/US06_PUBCOMB.seq:*
5: /cgn2_6/ptodata/1/pubpna/US07_NEW_PUB.seq:*
6: /cgn2_6/ptodata/1/pubpna/PCTUS_PUBCOMB.seq:*
7: /cgn2_6/ptodata/1/pubpna/US08_NEW_PUB.seq:*
8: /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq:*
9: /cgn2_6/ptodata/1/pubpna/US09A_PUBCOMB.seq:*
10: /cgn2_6/ptodata/1/pubpna/US09B_PUBCOMB.seq:*
11: /cgn2_6/ptodata/1/pubpna/US09C_PUBCOMB.seq:*
12: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq:*
13: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq2:*
14: /cgn2_6/ptodata/1/pubpna/US10A_PUBCOMB.seq:*
15: /cgn2_6/ptodata/1/pubpna/US10B_PUBCOMB.seq:*
16: /cgn2_6/ptodata/1/pubpna/US10C_PUBCOMB.seq:*
17: /cgn2_6/ptodata/1/pubpna/US10D_PUBCOMB.seq:*
18: /cgn2_6/ptodata/1/pubpna/US10E_PUBCOMB.seq:*
19: /cgn2_6/ptodata/1/pubpna/US10F_PUBCOMB.seq:*
20: /cgn2_6/ptodata/1/pubpna/US10G_PUBCOMB.seq:*
21: /cgn2_6/ptodata/1/pubpna/US10H_PUBCOMB.seq:*
22: /cgn2_6/ptodata/1/pubpna/US10I_PUBCOMB.seq:*
23: /cgn2_6/ptodata/1/pubpna/US10_NEW_PUB.seq:*
24: /cgn2_6/ptodata/1/pubpna/US10_NEW_PUB.seq2:*
25: /cgn2_6/ptodata/1/pubpna/US11A_PUBCOMB.seq:*
26: /cgn2_6/ptodata/1/pubpna/US11_NEW_PUB.seq:*
27: /cgn2_6/ptodata/1/pubpna/US60_NEW_PUB.seq:*
28: /cgn2_6/ptodata/1/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	1100	100.0	1100	11	US-09-774-954-16	Sequence 16, Appl
2	1100	100.0	1514	11	US-09-774-954-1	Sequence 1, Appli
3	1098.4	99.9	1300	11	US-09-774-954-4	Sequence 4, Appli
4	1098.4	99.9	11284	11	US-09-774-954-5	Sequence 5, Appli
5	1096.8	99.7	5218	16	US-10-301-822-168	Sequence 168, App
6	1096.8	99.7	5218	22	US-10-956-157-4752	Sequence 4752, Ap
7	1033.4	93.9	5009	11	US-09-774-954-7	Sequence 7, Appli
8	471.4	42.9	1400	22	US-10-956-157-9986	Sequence 9986, Ap
9	471.4	42.9	1506	22	US-10-956-157-4751	Sequence 4751, Ap
10	251.2	22.8	10331	10	US-09-764-891-6032	Sequence 6032, Ap
11	233	21.2	479	10	US-09-918-995-19437	Sequence 19437, A
12	221.8	20.2	477	10	US-09-918-995-22744	Sequence 22744, A
13	166.4	15.1	1209	26	US-11-097-143-18161	Sequence 18161, A
c 14	150.4	13.7	3264	26	US-11-097-143-18160	Sequence 18160, A
15	114.6	10.4	402	9	US-09-783-590-11501	Sequence 11501, A
c 16	108.6	9.9	3793	26	US-11-097-143-18082	Sequence 18082, A
c 17	71.8	6.5	3567	26	US-11-097-143-18169	Sequence 18169, A
18	60	5.5	60	10	US-09-908-975-17287	Sequence 17287, A
19	56	5.1	760	24	US-10-820-474A-236	Sequence 236, App
c 20	55.6	5.1	521	13	US-09-925-065A-602444	Sequence 602444,
21	42.4	3.9	1438	18	US-10-282-122A-19431	Sequence 19431, A
c 22	41.2	3.7	114793	16	US-10-148-806-3	Sequence 3, Appli
c 23	41.2	3.7	114793	24	US-10-859-792-3	Sequence 3, Appli
24	40.4	3.7	1212	21	US-10-425-115-114097	Sequence 114097,
25	40.4	3.7	5932	16	US-10-205-823-312	Sequence 312, App
26	40.4	3.7	5932	26	US-11-051-454-312	Sequence 312, App
27	40	3.6	40	11	US-09-774-954-10	Sequence 10, Appl
28	40	3.6	40	11	US-09-774-954-12	Sequence 12, Appl
29	40	3.6	2032	20	US-10-437-963-8098	Sequence 8098, Ap
c 30	39.2	3.6	618	13	US-09-925-065A-571171	Sequence 571171,
31	39	3.5	372	18	US-10-282-122A-31476	Sequence 31476, A
32	38.6	3.5	407	9	US-09-960-352-11732	Sequence 11732, A
33	38.6	3.5	2358	10	US-09-919-039-190	Sequence 190, App
34	38.6	3.5	2384	22	US-10-278-698-123	Sequence 123, App
35	38.6	3.5	2384	22	US-10-278-698-638	Sequence 638, App
36	38.6	3.5	2670	17	US-10-247-671-64	Sequence 64, Appl
37	38.4	3.5	532	10	US-09-918-995-13517	Sequence 13517, A
38	38.4	3.5	591	16	US-10-156-761-3550	Sequence 3550, Ap
39	38.4	3.5	6227	19	US-10-240-425-364	Sequence 364, App
40	38.4	3.5	6227	22	US-10-278-698-100	Sequence 100, App
41	38.4	3.5	6227	22	US-10-278-698-615	Sequence 615, App
42	38.4	3.5	6256	22	US-10-278-698-101	Sequence 101, App
43	38.4	3.5	6256	22	US-10-278-698-616	Sequence 616, App
44	38.4	3.5	9025608	16	US-10-156-761-1	Sequence 1, Appli
c 45	37.8	3.4	1234	18	US-10-369-493-39272	Sequence 39272, A

ALIGNMENTS

RESULT 1
US-09-774-954-16
; Sequence 16, Application US/09774954
; Publication No. US20040241645A1
; GENERAL INFORMATION:
; APPLICANT: Yang Wang, Michael W. Spellman
; TITLE OF INVENTION: O-Fucosyltransferase
; NUMBER OF SEQUENCES: 17
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genentech, Inc.
; STREET: 1 DNA Way
; CITY: South San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94080
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5 inch, 1.44 Mb floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: WinPatin (Genentech)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/774,954
; FILING DATE: 30-Jan-2001
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/978,741
; FILING DATE: 26-NOV-1997
; APPLICATION NUMBER: 08/792,498

```
;            FILING DATE: 31-JAN-1997
;       ATTORNEY/AGENT INFORMATION:
;            NAME: Svoboda, Craig G.
;            REGISTRATION NUMBER: 39,044
;            REFERENCE/DOCKET NUMBER: P1041P1
;       TELECOMMUNICATION INFORMATION:
;            TELEPHONE: 650/225-1489
;            TELEFAX: 650/952-9881
;   INFORMATION FOR SEQ ID NO: 16:
;       SEQUENCE CHARACTERISTICS:
;            LENGTH: 1100 base pairs
;            TYPE: Nucleic Acid
;            STRANDEDNESS: Single
;            TOPOLOGY: Linear
;       SEQUENCE DESCRIPTION: SEQ ID NO: 16:
US-09-774-954-16

  Query Match            100.0%;  Score 1100;  DB 11;  Length 1100;
  Best Local Similarity  100.0%;  Pred. No. 0;
  Matches 1100;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 ATGCCCGCGGGCTCCTGGGACCCGGCCGGTTACCTGCTCTACTGCCCCTGCATGGGGCGC 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGCCCGCGGGCTCCTGGGACCCGGCCGGTTACCTGCTCTACTGCCCCTGCATGGGGCGC 60

Qy         61 TTTGGGAACCAGGCCGATCACTTCTTGGGCTCTCTGGCATTTGCAAAGCTGCTAAACCGT 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 TTTGGGAACCAGGCCGATCACTTCTTGGGCTCTCTGGCATTTGCAAAGCTGCTAAACCGT 120

Qy        121 ACCTTGGCTGTCCCTCCTTGGATTGAGTACCAGCATCACAAGCCTCCTTTCACCAACCTC 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 ACCTTGGCTGTCCCTCCTTGGATTGAGTACCAGCATCACAAGCCTCCTTTCACCAACCTC 180

Qy        181 CATGTGTCCTACCAGAAGTACTTCAAGCTGGAGCCCCTCCAGGCTTACCATCGGGTCATC 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 CATGTGTCCTACCAGAAGTACTTCAAGCTGGAGCCCCTCCAGGCTTACCATCGGGTCATC 240

Qy        241 AGCTTGGAGGATTTCATGGAGAAGCTGGCACCCACCCACTGGCCCCCTGAGAAGCGGGTG 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 AGCTTGGAGGATTTCATGGAGAAGCTGGCACCCACCCACTGGCCCCCTGAGAAGCGGGTG 300

Qy        301 GCATACTGCTTTGAGGTGGCAGCCCAGCGAAGCCCAGATAAGAAGACGTGCCCCATGAAG 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 GCATACTGCTTTGAGGTGGCAGCCCAGCGAAGCCCAGATAAGAAGACGTGCCCCATGAAG 360

Qy        361 GAAGGAAACCCCTTTGGCCCATTCTGGGATCAGTTTCATGTGAGTTTCAACAAGTCGGAG 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 GAAGGAAACCCCTTTGGCCCATTCTGGGATCAGTTTCATGTGAGTTTCAACAAGTCGGAG 420

Qy        421 CTTTTTACAGGCATTTCCTTCAGTGCTTCCTACAGAGAACAATGGAGCCAGAGATTTTCT 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 CTTTTTACAGGCATTTCCTTCAGTGCTTCCTACAGAGAACAATGGAGCCAGAGATTTTCT 480

Qy        481 CCAAAGGAACATCCGGTGCTTGCCCTGCCAGGAGCCCCAGCCCAGTTCCCCGTCCTAGAA 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 CCAAAGGAACATCCGGTGCTTGCCCTGCCAGGAGCCCCAGCCCAGTTCCCCGTCCTAGAA 540

Qy        541 GAACACAGGCCACTACAGAAGTACATGGTATGGTCAGACGAAATGGTGAAGACGGGAGAG 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 GAACACAGGCCACTACAGAAGTACATGGTATGGTCAGACGAAATGGTGAAGACGGGAGAG 600

Qy        601 GCCCAGATTCATGCCCACCTTGTCCGGCCCTATGTGGGCATTCATCTGCGCATTGGCTCT 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 GCCCAGATTCATGCCCACCTTGTCCGGCCCTATGTGGGCATTCATCTGCGCATTGGCTCT 660

Qy        661 GACTGGAAGAACGCCTGTGCCATGCTGAAGGACGGGACTGCAGGCTCGCACTTCATGGCC 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 GACTGGAAGAACGCCTGTGCCATGCTGAAGGACGGGACTGCAGGCTCGCACTTCATGGCC 720

Qy        721 TCTCCGCAGTGTGTGGGCTACAGCCGCAGCACAGCGGCCCCCCTCACGATGACTATGTGC 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 TCTCCGCAGTGTGTGGGCTACAGCCGCAGCACAGCGGCCCCCCTCACGATGACTATGTGC 780

Qy        781 CTGCCTGACCTGAAGGAGATCCAGAGGGCTGTGAAGCTCTGGGTGAGGTCGCTGGATGCC 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 CTGCCTGACCTGAAGGAGATCCAGAGGGCTGTGAAGCTCTGGGTGAGGTCGCTGGATGCC 840

Qy        841 CAGTCGGTCTACGTTGCTACTGATTCCGAGAGTTATGTGCCTGAGCTCCAACAGCTCTTC 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 CAGTCGGTCTACGTTGCTACTGATTCCGAGAGTTATGTGCCTGAGCTCCAACAGCTCTTC 900

Qy        901 AAAGGGAAGGTGAAGGTGGTGAGCCTGAAGCCTGAGGTGGCCCAGGTCGACCTGTACATC 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 AAAGGGAAGGTGAAGGTGGTGAGCCTGAAGCCTGAGGTGGCCCAGGTCGACCTGTACATC 960

Qy        961 CTCGGCCAAGCCGACCACTTTATTGGCAACTGTGTCTCCTCCTTCACTGCCTTTGTGAAG 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 CTCGGCCAAGCCGACCACTTTATTGGCAACTGTGTCTCCTCCTTCACTGCCTTTGTGAAG 1020

Qy       1021 CGGGAGCGGGACCTCCAGGGGAGGCCGTCTTCTTTCTTCTCGGCATGGACAGGCCCCCTAAG 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1021 CGGGAGCGGGACCTCCAGGGGAGGCCGTCTTCTTTCTTCGGCATGGACAGGCCCCCTAAG 1080

Qy       1081 CTGCGGGACGAGTTCTGATT 1100
              ||||||||||||||||||||
Db       1081 CTGCGGGACGAGTTCTGATT 1100


RESULT 2
US-09-774-954-1
; Sequence 1, Application US/09774954
; Publication No. US20040241645A1
;   GENERAL INFORMATION:
;       APPLICANT: Yang Wang, Michael W. Spellman
;       TITLE OF INVENTION: O-Fucosyltransferase
;       NUMBER OF SEQUENCES: 17
;       CORRESPONDENCE ADDRESS:
;            ADDRESSEE: Genentech, Inc.
;            STREET: 1 DNA Way
;            CITY: South San Francisco
;            STATE: California
;            COUNTRY: USA
;            ZIP: 94080
;       COMPUTER READABLE FORM:
;            MEDIUM TYPE: 3.5 inch, 1.44 Mb floppy disk
;            COMPUTER: IBM PC compatible
;            OPERATING SYSTEM: PC-DOS/MS-DOS
;            SOFTWARE: WinPatin (Genentech)
;       CURRENT APPLICATION DATA:
;            APPLICATION NUMBER: US/09/774,954
;            FILING DATE: 30-Jan-2001
;            CLASSIFICATION: <Unknown>
;       PRIOR APPLICATION DATA:
;            APPLICATION NUMBER: US/08/978,741
;            FILING DATE: 26-NOV-1997
;            APPLICATION NUMBER: 08/792,498
;            FILING DATE: 31-JAN-1997
;       ATTORNEY/AGENT INFORMATION:
;            NAME: Svoboda, Craig G.
;            REGISTRATION NUMBER: 39,044
;            REFERENCE/DOCKET NUMBER: P1041P1
;       TELECOMMUNICATION INFORMATION:
;            TELEPHONE: 650/225-1489
;            TELEFAX: 650/952-9881
;   INFORMATION FOR SEQ ID NO: 1:
;       SEQUENCE CHARACTERISTICS:
;            LENGTH: 1514 base pairs
;            TYPE: Nucleic Acid
;            STRANDEDNESS: Single
;            TOPOLOGY: Linear
;       SEQUENCE DESCRIPTION: SEQ ID NO: 1:
US-09-774-954-1

  Query Match            100.0%;  Score 1100;  DB 11;  Length 1514;
  Best Local Similarity  100.0%;  Pred. No. 0;
  Matches 1100;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;
```

```
Qy          1 ATGCCCGCGGGCTCCTGGGACCCGGCCGGTTACCTGCTCTACTGCCCCTGCATGGGGCGC 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGCCCGCGGGCTCCTGGGACCCGGCCGGTTACCTGCTCTACTGCCCCTGCATGGGGCGC 60

Qy         61 TTTGGGAACCAGGCCGATCACTTCTTGGGCTCTCTGGCATTTGCAAAGCTGCTAAACCGT 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 TTTGGGAACCAGGCCGATCACTTCTTGGGCTCTCTGGCATTTGCAAAGCTGCTAAACCGT 120

Qy        121 ACCTTGGCTGTCCCTCCTTGGATTGAGTACCAGCATCACAAGCCTCCTTTCACCAACCTC 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 ACCTTGGCTGTCCCTCCTTGGATTGAGTACCAGCATCACAAGCCTCCTTTCACCAACCTC 180

Qy        181 CATGTGTCCTACCAGAAGTACTTCAAGCTGGAGCCCCTCCAGGCTTACCATCGGGTCATC 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 CATGTGTCCTACCAGAAGTACTTCAAGCTGGAGCCCCTCCAGGCTTACCATCGGGTCATC 240

Qy        241 AGCTTGGAGGATTTCATGGAGAAGCTGGCACCCACCCACTGGCCCCCTGAGAAGCGGGTG 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 AGCTTGGAGGATTTCATGGAGAAGCTGGCACCCACCCACTGGCCCCCTGAGAAGCGGGTG 300

Qy        301 GCATACTGCTTTGAGGTGGCAGCCCAGCGAAGCCCAGATAAGAAGACGTGCCCCATGAAG 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 GCATACTGCTTTGAGGTGGCAGCCCAGCGAAGCCCAGATAAGAAGACGTGCCCCATGAAG 360

Qy        361 GAAGGAAACCCCTTTGGCCCATTCTGGGATCAGTTTCATGTGAGTTTCAACAAGTCGGAG 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 GAAGGAAACCCCTTTGGCCCATTCTGGGATCAGTTTCATGTGAGTTTCAACAAGTCGGAG 420

Qy        421 CTTTTTACAGGCATTTCCTTCAGTGCTTCCTACAGAGAACAATGGAGCCAGAGATTTTCT 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 CTTTTTACAGGCATTTCCTTCAGTGCTTCCTACAGAGAACAATGGAGCCAGAGATTTTCT 480

Qy        481 CCAAAGGAACATCCGGTGCTTGCCCTGCCAGGAGCCCCAGCCCAGTTCCCCGTCCTAGAA 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 CCAAAGGAACATCCGGTGCTTGCCCTGCCAGGAGCCCCAGCCCAGTTCCCCGTCCTAGAA 540

Qy        541 GAACACAGGCCACTACAGAAGTACATGGTATGGTCAGACGAAATGGTGAAGACGGGAGAG 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 GAACACAGGCCACTACAGAAGTACATGGTATGGTCAGACGAAATGGTGAAGACGGGAGAG 600

Qy        601 GCCCAGATTCATGCCCACCTTGTCCGGCCCTATGTGGGCATTCATCTGCGCATTGGCTCT 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 GCCCAGATTCATGCCCACCTTGTCCGGCCCTATGTGGGCATTCATCTGCGCATTGGCTCT 660

Qy        661 GACTGGAAGAACGCCTGTGCCATGCTGAAGGACGGGACTGCAGGCTCGCACTTCATGGCC 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 GACTGGAAGAACGCCTGTGCCATGCTGAAGGACGGGACTGCAGGCTCGCACTTCATGGCC 720

Qy        721 TCTCCGCAGTGTGTGGGCTACAGCCGCAGCACAGCGGCCCCCCTCACGATGACTATGTGC 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 TCTCCGCAGTGTGTGGGCTACAGCCGCAGCACAGCGGCCCCCCTCACGATGACTATGTGC 780

Qy        781 CTGCCTGACCTGAAGGAGATCCAGAGGGCTGTGAAGCTCTGGGTGAGGTCGCTGGATGCC 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 CTGCCTGACCTGAAGGAGATCCAGAGGGCTGTGAAGCTCTGGGTGAGGTCGCTGGATGCC 840

Qy        841 CAGTCGGTCTACGTTGCTACTGATTCCGAGAGTTATGTGCCTGAGCTCCAACAGCTCTTC 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 CAGTCGGTCTACGTTGCTACTGATTCCGAGAGTTATGTGCCTGAGCTCCAACAGCTCTTC 900

Qy        901 AAAGGGAAGGTGAAGGTGGTGAGCCTGAAGCCTGAGGTGGCCCAGGTCGACCTGTACATC 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 AAAGGGAAGGTGAAGGTGGTGAGCCTGAAGCCTGAGGTGGCCCAGGTCGACCTGTACATC 960

Qy        961 CTCGGCCAAGCCGACCACTTTATTGGCAACTGTGTCTCCTCCTTCACTGCCTTTGTGAAG 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 CTCGGCCAAGCCGACCACTTTATTGGCAACTGTGTCTCCTCCTTCACTGCCTTTGTGAAG 1020

Qy       1021 CGGGAGCGGGACCTCCAGGGGAGGCCGTCTTCTTTCTTCGGCATGGACAGGCCCCCTAAG 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1021 CGGGAGCGGGACCTCCAGGGGAGGCCGTCTTCTTTCTTCGGCATGGACAGGCCCCCTAAG 1080
```

```
Qy       1081 CTGCGGGACGAGTTCTGATT 1100
              ||||||||||||||||||||
Db       1081 CTGCGGGACGAGTTCTGATT 1100


RESULT 3
US-09-774-954-4
; Sequence 4, Application US/09774954
; Publication No. US20040241645A1
;  GENERAL INFORMATION:
;       APPLICANT: Yang Wang, Michael W. Spellman
;       TITLE OF INVENTION: O-Fucosyltransferase
;       NUMBER OF SEQUENCES: 17
;       CORRESPONDENCE ADDRESS:
;            ADDRESSEE: Genentech, Inc.
;            STREET: 1 DNA Way
;            CITY: South San Francisco
;            STATE: California
;            COUNTRY: USA
;            ZIP: 94080
;       COMPUTER READABLE FORM:
;            MEDIUM TYPE: 3.5 inch, 1.44 Mb floppy disk
;            COMPUTER: IBM PC compatible
;            OPERATING SYSTEM: PC-DOS/MS-DOS
;            SOFTWARE: WinPatin (Genentech)
;       CURRENT APPLICATION DATA:
;            APPLICATION NUMBER: US/09/774,954
;            FILING DATE: 30-Jan-2001
;            CLASSIFICATION: <Unknown>
;       PRIOR APPLICATION DATA:
;            APPLICATION NUMBER: US/08/978,741
;            FILING DATE: 26-NOV-1997
;            APPLICATION NUMBER: 08/792,498
;            FILING DATE: 31-JAN-1997
;       ATTORNEY/AGENT INFORMATION:
;            NAME: Svoboda, Craig G.
;            REGISTRATION NUMBER: 39,044
;            REFERENCE/DOCKET NUMBER: P1041P1
;       TELECOMMUNICATION INFORMATION:
;            TELEPHONE: 650/225-1489
;            TELEFAX: 650/952-9881
;  INFORMATION FOR SEQ ID NO: 4:
;       SEQUENCE CHARACTERISTICS:
;            LENGTH: 1300 base pairs
;            TYPE: Nucleic Acid
;            STRANDEDNESS: Single
;            TOPOLOGY: Linear
;       SEQUENCE DESCRIPTION: SEQ ID NO: 4:
US-09-774-954-4

  Query Match             99.9%;  Score 1098.4;  DB 11;  Length 1300;
  Best Local Similarity   99.9%;  Pred. No. 0;
  Matches 1099;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;
```

```
Qy          1 ATGCCCGCGGGCTCCTGGGACCCGGCCGGTTACCTGCTCTACTGCCCCTGCATGGGGCGC 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        136 ATGCCCGCGGGCTCCTGGGACCCGGCCGGTTACCTGCTCTACTGCCCCTGCATGGGGCGC 195

Qy         61 TTTGGGAACCAGGCCGATCACTTCTTGGGCTCTCTGGCATTTGCAAAGCTGCTAAACCGT 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        196 TTTGGGAACCAGGCCGATCACTTCTTGGGCTCTCTGGCATTTGCAAAGCTGCTAAACCGT 255

Qy        121 ACCTTGGCTGTCCCTCCTTGGATTGAGTACCAGCATCACAAGCCTCCTTTCACCAACCTC 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        256 ACCTTGGCTGTCCCTCCTTGGATTGAGTACCAGCATCACAAGCCTCCTTTCACCAACCTC 315

Qy        181 CATGTGTCCTACCAGAAGTACTTCAAGCTGGAGCCCCTCCAGGCTTACCATCGGGTCATC 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        316 CATGTGTCCTACCAGAAGTACTTCAAGCTGGAGCCCCTCCAGGCTTACCATCGGGTCATC 375

Qy        241 AGCTTGGAGGATTTCATGGAGAAGCTGGCACCCACCCACTGGCCCCCTGAGAAGCGGGTG 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        376 AGCTTGGAGGATTTCATGGAGAAGCTGGCACCCACCCACTGGCCCCCTGAGAAGCGGGTG 435
```

```
Qy        301 GCATACTGCTTTGAGGTGGCAGCCCAGCGAAGCCCAGATAAGAAGACGTGCCCCATGAAG 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        436 GCATACTGCTTTGAGGTGGCAGCCCAGCGAAGCCCAGATAAGAAGACGTGCCCCATGAAG 495

Qy        361 GAAGGAAACCCCTTTGGCCCATTCTGGGATCAGTTTCATGTGAGTTTCAACAAGTCGGAG 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        496 GAAGGAAACCCCTTTGGCCCATTCTGGGATCAGTTTCATGTGAGTTTCAACAAGTCGGAG 555

Qy        421 CTTTTTACAGGCATTTCCTTCAGTGCTTCCTACAGAGAACAATGGAGCCAGAGATTTTCT 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        556 CTTTTTACAGGCATTTCCTTCAGTGCTTCCTACAGAGAACAATGGAGCCAGAGATTTTCT 615

Qy        481 CCAAAGGAACATCCGGTGCTTGCCCTGCCAGGAGCCCCAGCCCAGTTCCCCGTCCTAGAA 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        616 CCAAAGGAACATCCGGTGCTTGCCCTGCCAGGAGCCCCAGCCCAGTTCCCCGTCCTAGAG 675

Qy        541 GAACACAGGCCACTACAGAAGTACATGGTATGGTCAGACGAAATGGTGAAGACGGGAGAG 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        676 GAACACAGGCCACTACAGAAGTACATGGTATGGTCAGACGAAATGGTGAAGACGGGAGAG 735

Qy        601 GCCCAGATTCATGCCCACCTTGTCCGGCCCTATGTGGGCATTCATCTGCGCATTGGCTCT 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        736 GCCCAGATTCATGCCCACCTTGTCCGGCCCTATGTGGGCATTCATCTGCGCATTGGCTCT 795

Qy        661 GACTGGAAGAACGCCTGTGCCATGCTGAAGGACGGGACTGCAGGCTCGCACTTCATGGCC 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        796 GACTGGAAGAACGCCTGTGCCATGCTGAAGGACGGGACTGCAGGCTCGCACTTCATGGCC 855

Qy        721 TCTCCGCAGTGTGTGGGCTACAGCCGCAGCACAGCGGCCCCCCTCACGATGACTATGTGC 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        856 TCTCCGCAGTGTGTGGGCTACAGCCGCAGCACAGCGGCCCCCCTCACGATGACTATGTGC 915

Qy        781 CTGCCTGACCTGAAGGAGATCCAGAGGGCTGTGAAGCTCTGGGTGAGGTCGCTGGATGCC 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        916 CTGCCTGACCTGAAGGAGATCCAGAGGGCTGTGAAGCTCTGGGTGAGGTCGCTGGATGCC 975

Qy        841 CAGTCGGTCTACGTTGCTACTGATTCCGAGAGTTATGTGCCTGAGCTCCAACAGCTCTTC 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        976 CAGTCGGTCTACGTTGCTACTGATTCCGAGAGTTATGTGCCTGAGCTCCAACAGCTCTTC 1035

Qy        901 AAAGGGAAGGTGAAGGTGGTGAGCCTGAAGCCTGAGGTGGCCCAGGTCGACCTGTACATC 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1036 AAAGGGAAGGTGAAGGTGGTGAGCCTGAAGCCTGAGGTGGCCCAGGTCGACCTGTACATC 1095

Qy        961 CTCGGCCAAGCCGACCACTTTATTGGCAACTGTGTCTCCTCCTTCACTGCCTTTGTGAAG 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1096 CTCGGCCAAGCCGACCACTTTATTGGCAACTGTGTCTCCTCCTTCACTGCCTTTGTGAAG 1155

Qy       1021 CGGGAGCGGGACCTCCAGGGGAGGCCGTCTTCTTTCTTCGGCATGGACAGGCCCCCTAAG 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1156 CGGGAGCGGGACCTCCAGGGGAGGCCGTCTTCTTTCTTCGGCATGGACAGGCCCCCTAAG 1215

Qy       1081 CTGCGGGACGAGTTCTGATT 1100
              ||||||||||||||||||||
Db       1216 CTGCGGGACGAGTTCTGATT 1235


RESULT 4
US-09-774-954-5
; Sequence 5, Application US/09774954
; Publication No. US20040241645A1
;   GENERAL INFORMATION:
;        APPLICANT: Yang Wang, Michael W. Spellman
;        TITLE OF INVENTION: O-Fucosyltransferase
;        NUMBER OF SEQUENCES: 17
;        CORRESPONDENCE ADDRESS:
;             ADDRESSEE: Genentech, Inc.
;             STREET: 1 DNA Way
;             CITY: South San Francisco
;             STATE: California
;             COUNTRY: USA
;             ZIP: 94080
;        COMPUTER READABLE FORM:
;             MEDIUM TYPE: 3.5 inch, 1.44 Mb floppy disk
;             COMPUTER: IBM PC compatible
;             OPERATING SYSTEM: PC-DOS/MS-DOS
;             SOFTWARE: WinPatin (Genentech)
;        CURRENT APPLICATION DATA:
;             APPLICATION NUMBER: US/09/774,954
;             FILING DATE: 30-Jan-2001
;             CLASSIFICATION: <Unknown>
;        PRIOR APPLICATION DATA:
;             APPLICATION NUMBER: US/08/978,741
;             FILING DATE: 26-NOV-1997
;             APPLICATION NUMBER: 08/792,498
;             FILING DATE: 31-JAN-1997
;        ATTORNEY/AGENT INFORMATION:
;             NAME: Svoboda, Craig G.
;             REGISTRATION NUMBER: 39,044
;             REFERENCE/DOCKET NUMBER: P1041P1
;        TELECOMMUNICATION INFORMATION:
;             TELEPHONE: 650/225-1489
;             TELEFAX: 650/952-9881
;   INFORMATION FOR SEQ ID NO: 5:
;        SEQUENCE CHARACTERISTICS:
;             LENGTH: 11284 base pairs
;             TYPE: Nucleic Acid
;             STRANDEDNESS: Single
;             TOPOLOGY: Linear
;        SEQUENCE DESCRIPTION: SEQ ID NO: 5:
US-09-774-954-5

  Query Match             99.9%;  Score 1098.4;  DB 11;  Length 11284;
  Best Local Similarity   99.9%;  Pred. No. 0;
  Matches 1099;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 ATGCCCGCGGGCTCCTGGGACCCGGCCGGTTACCTGCTCTACTGCCCCTGCATGGGGCGC 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4236 ATGCCCGCGGGCTCCTGGGACCCGGCCGGTTACCTGCTCTACTGCCCCTGCATGGGGCGC 4295

Qy         61 TTTGGGAACCAGGCCGATCACTTCTTGGGCTCTCTGGCATTTGCAAAGCTGCTAAACCGT 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4296 TTTGGGAACCAGGCCGATCACTTCTTGGGCTCTCTGGCATTTGCAAAGCTGCTAAACCGT 4355

Qy        121 ACCTTGGCTGTCCCTCCTTGGATTGAGTACCAGCATCACAAGCCTCCTTTCACCAACCTC 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4356 ACCTTGGCTGTCCCTCCTTGGATTGAGTACCAGCATCACAAGCCTCCTTTCACCAACCTC 4415

Qy        181 CATGTGTCCTACCAGAAGTACTTCAAGCTGGAGCCCCTCCAGGCTTACCATCGGGTCATC 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4416 CATGTGTCCTACCAGAAGTACTTCAAGCTGGAGCCCCTCCAGGCTTACCATCGGGTCATC 4475

Qy        241 AGCTTGGAGGATTTCATGGAGAAGCTGGCACCCACCCACTGGCCCCCTGAGAAGCGGGTG 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4476 AGCTTGGAGGATTTCATGGAGAAGCTGGCACCCACCCACTGGCCCCCTGAGAAGCGGGTG 4535

Qy        301 GCATACTGCTTTGAGGTGGCAGCCCAGCGAAGCCCAGATAAGAAGACGTGCCCCATGAAG 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4536 GCATACTGCTTTGAGGTGGCAGCCCAGCGAAGCCCAGATAAGAAGACGTGCCCCATGAAG 4595

Qy        361 GAAGGAAACCCCTTTGGCCCATTCTGGGATCAGTTTCATGTGAGTTTCAACAAGTCGGAG 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4596 GAAGGAAACCCCTTTGGCCCATTCTGGGATCAGTTTCATGTGAGTTTCAACAAGTCGGAG 4655

Qy        421 CTTTTTACAGGCATTTCCTTCAGTGCTTCCTACAGAGAACAATGGAGCCAGAGATTTTCT 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4656 CTTTTTACAGGCATTTCCTTCAGTGCTTCCTACAGAGAACAATGGAGCCAGAGATTTTCT 4715

Qy        481 CCAAAGGAACATCCGGTGCTTGCCCTGCCAGGAGCCCCAGCCCAGTTCCCCGTCCTAGAA 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4716 CCAAAGGAACATCCGGTGCTTGCCCTGCCAGGAGCCCCAGCCCAGTTCCCCGTCCTAGAG 4775

Qy        541 GAACACAGGCCACTACAGAAGTACATGGTATGGTCAGACGAAATGGTGAAGACGGGAGAG 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4776 GAACACAGGCCACTACAGAAGTACATGGTATGGTCAGACGAAATGGTGAAGACGGGAGAG 4835
```

```
Qy        601 GCCCAGATTCATGCCCACCTTGTCCGGCCCTATGTGGGCATTCATCTGCGCATTGGCTCT 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4836 GCCCAGATTCATGCCCACCTTGTCCGGCCCTATGTGGGCATTCATCTGCGCATTGGCTCT 4895

Qy        661 GACTGGAAGAACGCCTGTGCCATGCTGAAGGACGGGACTGCAGGCTCGCACTTCATGGCC 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4896 GACTGGAAGAACGCCTGTGCCATGCTGAAGGACGGGACTGCAGGCTCGCACTTCATGGCC 4955

Qy        721 TCTCCGCAGTGTGTGGGCTACAGCCGCAGCACAGCGGCCCCCCTCACGATGACTATGTGC 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4956 TCTCCGCAGTGTGTGGGCTACAGCCGCAGCACAGCGGCCCCCCTCACGATGACTATGTGC 5015

Qy        781 CTGCCTGACCTGAAGGAGATCCAGAGGGCTGTGAAGCTCTGGGTGAGGTCGCTGGATGCC 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5016 CTGCCTGACCTGAAGGAGATCCAGAGGGCTGTGAAGCTCTGGGTGAGGTCGCTGGATGCC 5075

Qy        841 CAGTCGGTCTACGTTGCTACTGATTCCGAGAGTTATGTGCCTGAGCTCCAACAGCTCTTC 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5076 CAGTCGGTCTACGTTGCTACTGATTCCGAGAGTTATGTGCCTGAGCTCCAACAGCTCTTC 5135

Qy        901 AAAGGGAAGGTGAAGGTGGTGAGCCTGAAGCCTGAGGTGGCCCAGGTCGACCTGTACATC 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5136 AAAGGGAAGGTGAAGGTGGTGAGCCTGAAGCCTGAGGTGGCCCAGGTCGACCTGTACATC 5195

Qy        961 CTCGGCCAAGCCGACCACTTTATTGGCAACTGTGTCTCCTCCTTCACTGCCTTTGTGAAG 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5196 CTCGGCCAAGCCGACCACTTTATTGGCAACTGTGTCTCCTCCTTCACTGCCTTTGTGAAG 5255

Qy       1021 CGGGAGCGGGACCTCCAGGGGAGGCCGTCTTCTTTCTTCGGCATGGACAGGCCCCCTAAG 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5256 CGGGAGCGGGACCTCCAGGGGAGGCCGTCTTCTTTCTTCGGCATGGACAGGCCCCCTAAG 5315

Qy       1081 CTGCGGGACGAGTTCTGATT 1100
              ||||||||||||||||||||
Db       5316 CTGCGGGACGAGTTCTGATT 5335
```

```
RESULT 5
US-10-301-822-168
; Sequence 168, Application US/10301822
; Publication No. US20030148410A1
; GENERAL INFORMATION:
;  APPLICANT: Millennium Pharmaceuticals, Inc.
;  APPLICANT:  Berger, Allison
;  APPLICANT:  Guillemette, Tracy L.
;  APPLICANT:  Kamatkar, Shubhangi
;  APPLICANT:  Schlegel, Robert
;  APPLICANT:  Monahan, John E.
;  APPLICANT:  Thibodeau, Stephen N.
;  APPLICANT:  Burgart, Lawrence J.
;  TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND
;  TITLE OF INVENTION:  METHODS FOR IDENTIFICATION, ASSESSMENT, PREVENTION, AND
;  TITLE OF INVENTION:  THERAPY OF COLON CANCER
;  FILE REFERENCE: MPM01-029P2RNM
;  CURRENT APPLICATION NUMBER: US/10/301,822
;  CURRENT FILING DATE:  2002-11-21
;  PRIOR APPLICATION NUMBER: US 60/339,971
;  PRIOR FILING DATE: 2001-12-10
;  PRIOR APPLICATION NUMBER: US 60/361,978
;  PRIOR FILING DATE: 2002-03-05
;  PRIOR APPLICATION NUMBER: US 60/381,988
;  PRIOR FILING DATE: 2002-05-20
;  NUMBER OF SEQ ID NOS: 228
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 168
;   LENGTH: 5218
;   TYPE: DNA
;   ORGANISM: Homo Sapiens
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (50)...(1216)
US-10-301-822-168
```

```
Query Match           99.7%;  Score 1096.8;  DB 16;  Length 5218;
Best Local Similarity  99.8%;  Pred. No. 0;
Matches 1098;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy          1 ATGCCCGCGGGCTCCTGGGACCCGGCCGGTTACCTGCTCTACTGCCCCTGCATGGGGCGC 60
              ||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        119 ATGCCTGCGGGCTCCTGGGACCCGGCCGGTTACCTGCTCTACTGCCCCTGCATGGGGCGC 178

Qy         61 TTTGGGAACCAGGCCGATCACTTCTTGGGCTCTCTGGCATTTGCAAAGCTGCTAAACCGT 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        179 TTTGGGAACCAGGCCGATCACTTCTTGGGCTCTCTGGCATTTGCAAAGCTGCTAAACCGT 238

Qy        121 ACCTTGGCTGTCCCTCCTTGGATTGAGTACCAGCATCACAAGCCTCCTTTCACCAACCTC 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        239 ACCTTGGCTGTCCCTCCTTGGATTGAGTACCAGCATCACAAGCCTCCTTTCACCAACCTC 298

Qy        181 CATGTGTCCTACCAGAAGTACTTCAAGCTGGAGCCCCTCCAGGCTTACCATCGGGTCATC 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        299 CATGTGTCCTACCAGAAGTACTTCAAGCTGGAGCCCCTCCAGGCTTACCATCGGGTCATC 358

Qy        241 AGCTTGGAGGATTTCATGGAGAAGCTGGCACCCACCCACTGGCCCCCTGAGAAGCGGGTG 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        359 AGCTTGGAGGATTTCATGGAGAAGCTGGCACCCACCCACTGGCCCCCTGAGAAGCGGGTG 418

Qy        301 GCATACTGCTTTGAGGTGGCAGCCCAGCGAAGCCCAGATAAGAAGACGTGCCCCATGAAG 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        419 GCATACTGCTTTGAGGTGGCAGCCCAGCGAAGCCCAGATAAGAAGACGTGCCCCATGAAG 478

Qy        361 GAAGGAAACCCCTTTGGCCCATTCTGGGATCAGTTTCATGTGAGTTTCAACAAGTCGGAG 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        479 GAAGGAAACCCCTTTGGCCCATTCTGGGATCAGTTTCATGTGAGTTTCAACAAGTCGGAG 538

Qy        421 CTTTTTACAGGCATTTCCTTCAGTGCTTCCTACAGAGAACAATGGAGCCAGAGATTTTCT 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        539 CTTTTTACAGGCATTTCCTTCAGTGCTTCCTACAGAGAACAATGGAGCCAGAGATTTTCT 598

Qy        481 CCAAAGGAACATCCGGTGCTTGCCCTGCCAGGAGCCCCAGCCCAGTTCCCCGTCCTAGAA 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        599 CCAAAGGAACATCCGGTGCTTGCCCTGCCAGGAGCCCCAGCCCAGTTCCCCGTCCTAGAG 658

Qy        541 GAACACAGGCCACTACAGAAGTACATGGTATGGTCAGACGAAATGGTGAAGACGGGAGAG 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        659 GAACACAGGCCACTACAGAAGTACATGGTATGGTCAGACGAAATGGTGAAGACGGGAGAG 718

Qy        601 GCCCAGATTCATGCCCACCTTGTCCGGCCCTATGTGGGCATTCATCTGCGCATTGGCTCT 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        719 GCCCAGATTCATGCCCACCTTGTCCGGCCCTATGTGGGCATTCATCTGCGCATTGGCTCT 778

Qy        661 GACTGGAAGAACGCCTGTGCCATGCTGAAGGACGGGACTGCAGGCTCGCACTTCATGGCC 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        779 GACTGGAAGAACGCCTGTGCCATGCTGAAGGACGGGACTGCAGGCTCGCACTTCATGGCC 838

Qy        721 TCTCCGCAGTGTGTGGGCTACAGCCGCAGCACAGCGGCCCCCCTCACGATGACTATGTGC 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        839 TCTCCGCAGTGTGTGGGCTACAGCCGCAGCACAGCGGCCCCCCTCACGATGACTATGTGC 898

Qy        781 CTGCCTGACCTGAAGGAGATCCAGAGGGCTGTGAAGCTCTGGGTGAGGTCGCTGGATGCC 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        899 CTGCCTGACCTGAAGGAGATCCAGAGGGCTGTGAAGCTCTGGGTGAGGTCGCTGGATGCC 958

Qy        841 CAGTCGGTCTACGTTGCTACTGATTCCGAGAGTTATGTGCCTGAGCTCCAACAGCTCTTC 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        959 CAGTCGGTCTACGTTGCTACTGATTCCGAGAGTTATGTGCCTGAGCTCCAACAGCTCTTC 1018

Qy        901 AAAGGGAAGGTGAAGGTGGTGAGCCTGAAGCCTGAGGTGGCCCAGGTCGACCTGTACATC 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1019 AAAGGGAAGGTGAAGGTGGTGAGCCTGAAGCCTGAGGTGGCCCAGGTCGACCTGTACATC 1078

Qy        961 CTCGGCCAAGCCGACCACTTTATTGGCAACTGTGTCTCCTCCTTCACTGCCTTTGTGAAG 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1079 CTCGGCCAAGCCGACCACTTTATTGGCAACTGTGTCTCCTCCTTCACTGCCTTTGTGAAG 1138
```

```
Qy       1021 CGGGAGCGGGACCTCCAGGGGAGGCCGTCTTCTTTCTTCGGCATGGACAGGCCCCCTAAG 1080
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1139 CGGGAGCGGGACCTCCAGGGGAGGCCGTCTTCTTTCTTCGGCATGGACAGGCCCCCTAAG 1198

Qy       1081 CTGCGGGACGAGTTCTGATT 1100
              ||||||||||||||||||||
Db       1199 CTGCGGGACGAGTTCTGATT 1218


RESULT 6
US-10-956-157-4752
; Sequence 4752, Application US/10956157
; Publication No. US20050118625A1
; GENERAL INFORMATION:
;  APPLICANT: Wyeth
;  APPLICANT:  Mounts, William
;  TITLE OF INVENTION: NUCLEIC ACID ARRAYS FOR DETECTING GENE EXPRESSION ASSOCIATED WITH
;  TITLE OF INVENTION:  HUMAN OSTEOARTHRITIS AND HUMAN PROTEASES
;  FILE REFERENCE: 031896-043000 (AM 101081)
;  CURRENT APPLICATION NUMBER: US/10/956,157
;  CURRENT FILING DATE:  2004-10-04
;  NUMBER OF SEQ ID NOS: 319805
;  SOFTWARE: PatentIn version 3.2
; SEQ ID NO 4752
;   LENGTH: 5218
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-10-956-157-4752

  Query Match             99.7%;  Score 1096.8;  DB 22;  Length 5218;
  Best Local Similarity   99.8%;  Pred. No. 0;
  Matches 1098;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy          1 ATGCCCGCGGGCTCCTGGGACCCGGCCGGTTACCTGCTCTACTGCCCCTGCATGGGGCGC 60
              ||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        119 ATGCCTGCGGGCTCCTGGGACCCGGCCGGTTACCTGCTCTACTGCCCCTGCATGGGGCGC 178

Qy         61 TTTGGGAACCAGGCCGATCACTTCTTGGGCTCTCTGGCATTTGCAAAGCTGCTAAACCGT 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        179 TTTGGGAACCAGGCCGATCACTTCTTGGGCTCTCTGGCATTTGCAAAGCTGCTAAACCGT 238

Qy        121 ACCTTGGCTGTCCCTCCTTGGATTGAGTACCAGCATCACAAGCCTCCTTTCACCAACCTC 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        239 ACCTTGGCTGTCCCTCCTTGGATTGAGTACCAGCATCACAAGCCTCCTTTCACCAACCTC 298

Qy        181 CATGTGTCCTACCAGAAGTACTTCAAGCTGGAGCCCCTCCAGGCTTACCATCGGGTCATC 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        299 CATGTGTCCTACCAGAAGTACTTCAAGCTGGAGCCCCTCCAGGCTTACCATCGGGTCATC 358

Qy        241 AGCTTGGAGGATTTCATGGAGAAGCTGGCACCCACCCACTGGCCCCCTGAGAAGCGGGTG 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        359 AGCTTGGAGGATTTCATGGAGAAGCTGGCACCCACCCACTGGCCCCCTGAGAAGCGGGTG 418

Qy        301 GCATACTGCTTTGAGGTGGCAGCCCAGCGAAGCCCAGATAAGAAGACGTGCCCCATGAAG 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        419 GCATACTGCTTTGAGGTGGCAGCCCAGCGAAGCCCAGATAAGAAGACGTGCCCCATGAAG 478

Qy        361 GAAGGAAACCCCTTTGGCCCATTCTGGGATCAGTTTCATGTGAGTTTCAACAAGTCGGAG 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        479 GAAGGAAACCCCTTTGGCCCATTCTGGGATCAGTTTCATGTGAGTTTCAACAAGTCGGAG 538

Qy        421 CTTTTTACAGGCATTTCCTTCAGTGCTTCCTACAGAGAACAATGGAGCCAGAGATTTTCT 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        539 CTTTTTACAGGCATTTCCTTCAGTGCTTCCTACAGAGAACAATGGAGCCAGAGATTTTCT 598

Qy        481 CCAAAGGAACATCCGGTGCTTGCCCTGCCAGGAGCCCCAGCCCAGTTCCCCGTCCTAGAA 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        599 CCAAAGGAACATCCGGTGCTTGCCCTGCCAGGAGCCCCAGCCCAGTTCCCCGTCCTAGAG 658

Qy        541 GAACACAGGCCACTACAGAAGTACATGGTATGGTCAGACGAAATGGTGAAGACGGGAGAG 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        659 GAACACAGGCCACTACAGAAGTACATGGTATGGTCAGACGAAATGGTGAAGACGGGAGAG 718

Qy        601 GCCCAGATTCATGCCCACCTTGTCCGGCCCTATGTGGGCATTCATCTGCGCATTGGCTCT 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        719 GCCCAGATTCATGCCCACCTTGTCCGGCCCTATGTGGGCATTCATCTGCGCATTGGCTCT 778

Qy        661 GACTGGAAGAACGCCTGTGCCATGCTGAAGGACGGGACTGCAGGCTCGCACTTCATGGCC 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        779 GACTGGAAGAACGCCTGTGCCATGCTGAAGGACGGGACTGCAGGCTCGCACTTCATGGCC 838

Qy        721 TCTCCGCAGTGTGTGGGCTACAGCCGCAGCACAGCGGCCCCCCTCACGATGACTATGTGC 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        839 TCTCCGCAGTGTGTGGGCTACAGCCGCAGCACAGCGGCCCCCCTCACGATGACTATGTGC 898

Qy        781 CTGCCTGACCTGAAGGAGATCCAGAGGGCTGTGAAGCTCTGGGTGAGGTCGCTGGATGCC 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        899 CTGCCTGACCTGAAGGAGATCCAGAGGGCTGTGAAGCTCTGGGTGAGGTCGCTGGATGCC 958

Qy        841 CAGTCGGTCTACGTTGCTACTGATTCCGAGAGTTATGTGCCTGAGCTCCAACAGCTCTTC 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        959 CAGTCGGTCTACGTTGCTACTGATTCCGAGAGTTATGTGCCTGAGCTCCAACAGCTCTTC 1018

Qy        901 AAAGGGAAGGTGAAGGTGGTGAGCCTGAAGCCTGAGGTGGCCCAGGTCGACCTGTACATC 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1019 AAAGGGAAGGTGAAGGTGGTGAGCCTGAAGCCTGAGGTGGCCCAGGTCGACCTGTACATC 1078

Qy        961 CTCGGCCAAGCCGACCACTTTATTGGCAACTGTGTCTCCTCCTTCACTGCCTTTGTGAAG 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1079 CTCGGCCAAGCCGACCACTTTATTGGCAACTGTGTCTCCTCCTTCACTGCCTTTGTGAAG 1138

Qy       1021 CGGGAGCGGGACCTCCAGGGGAGGCCGTCTTCTTTCTTCGGCATGGACAGGCCCCCTAAG 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1139 CGGGAGCGGGACCTCCAGGGGAGGCCGTCTTCTTTCTTCGGCATGGACAGGCCCCCTAAG 1198

Qy       1081 CTGCGGGACGAGTTCTGATT 1100
              ||||||||||||||||||||
Db       1199 CTGCGGGACGAGTTCTGATT 1218


RESULT 7
US-09-774-954-7
; Sequence 7, Application US/09774954
; Publication No. US20040241645A1
;   GENERAL INFORMATION:
;        APPLICANT: Yang Wang, Michael W. Spellman
;        TITLE OF INVENTION: O-Fucosyltransferase
;        NUMBER OF SEQUENCES: 17
;        CORRESPONDENCE ADDRESS:
;             ADDRESSEE: Genentech, Inc.
;             STREET: 1 DNA Way
;             CITY: South San Francisco
;             STATE: California
;             COUNTRY: USA
;             ZIP: 94080
;        COMPUTER READABLE FORM:
;             MEDIUM TYPE: 3.5 inch, 1.44 Mb floppy disk
;             COMPUTER: IBM PC compatible
;             OPERATING SYSTEM: PC-DOS/MS-DOS
;             SOFTWARE: WinPatin (Genentech)
;        CURRENT APPLICATION DATA:
;             APPLICATION NUMBER: US/09/774,954
;             FILING DATE: 30-Jan-2001
;             CLASSIFICATION: <Unknown>
;        PRIOR APPLICATION DATA:
;             APPLICATION NUMBER: US/08/978,741
;             FILING DATE: 26-NOV-1997
;             APPLICATION NUMBER: 08/792,498
;             FILING DATE: 31-JAN-1997
;        ATTORNEY/AGENT INFORMATION:
;             NAME: Svoboda, Craig G.
;             REGISTRATION NUMBER: 39,044
;             REFERENCE/DOCKET NUMBER: P1041P1
;        TELECOMMUNICATION INFORMATION:
;             TELEPHONE: 650/225-1489
;             TELEFAX: 650/952-9881
```

```
;   INFORMATION FOR SEQ ID NO: 7:
;        SEQUENCE CHARACTERISTICS:
;             LENGTH: 5009 base pairs
;             TYPE: Nucleic Acid
;             STRANDEDNESS: Single
;             TOPOLOGY: Linear
;        SEQUENCE DESCRIPTION: SEQ ID NO: 7:
US-09-774-954-7

  Query Match             93.9%;  Score 1033.4;  DB 11;  Length 5009;
  Best Local Similarity   99.9%;  Pred. No. 4.1e-314;
  Matches 1034;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          66 GAACCAGGCCGATCACTTCTTGGGCTCTCTGGCATTTGCAAAGCTGCTAAACCGTACCTT 125
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db           1 GAACCAGGCCGATCACTTCTTGGGCTCTCTGGCATTTGCAAAGCTGCTAAACCGTACCTT 60

Qy         126 GGCTGTCCCTCCTTGGATTGAGTACCAGCATCACAAGCCTCCTTTCACCAACCTCCATGT 185
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          61 GGCTGTCCCTCCTTGGATTGAGTACCAGCATCACAAGCCTCCTTTCACCAACCTCCATGT 120

Qy         186 GTCCTACCAGAAGTACTTCAAGCTGGAGCCCCTCCAGGCTTACCATCGGGTCATCAGCTT 245
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         121 GTCCTACCAGAAGTACTTCAAGCTGGAGCCCCTCCAGGCTTACCATCGGGTCATCAGCTT 180

Qy         246 GGAGGATTTCATGGAGAAGCTGGCACCCACCCACTGGCCCCCTGAGAAGCGGGTGGCATA 305
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         181 GGAGGATTTCATGGAGAAGCTGGCACCCACCCACTGGCCCCCTGAGAAGCGGGTGGCATA 240

Qy         306 CTGCTTTGAGGTGGCAGCCCAGCGAAGCCCAGATAAGAAGACGTGCCCCATGAAGGAAGG 365
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         241 CTGCTTTGAGGTGGCAGCCCAGCGAAGCCCAGATAAGAAGACGTGCCCCATGAAGGAAGG 300

Qy         366 AAACCCCTTTGGCCCATTCTGGGATCAGTTTCATGTGAGTTTCAACAAGTCGGAGCTTTT 425
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         301 AAACCCCTTTGGCCCATTCTGGGATCAGTTTCATGTGAGTTTCAACAAGTCGGAGCTTTT 360

Qy         426 TACAGGCATTTCCTTCAGTGCTTCCTACAGAGAACAATGGAGCCAGAGATTTTCTCCAAA 485
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         361 TACAGGCATTTCCTTCAGTGCTTCCTACAGAGAACAATGGAGCCAGAGATTTTCTCCAAA 420

Qy         486 GGAACATCCGGTGCTTGCCCTGCCAGGAGCCCCAGCCCAGTTCCCCGTCCTAGAAGAACA 545
               |||||||||||||||||||||||||||||||||||||||||||||||||| |||||||||
Db         421 GGAACATCCGGTGCTTGCCCTGCCAGGAGCCCCAGCCCAGTTCCCCGTCCTAGAGGAACA 480

Qy         546 CAGGCCACTACAGAAGTACATGGTATGGTCAGACGAAATGGTGAAGACGGGAGAGGCCCA 605
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         481 CAGGCCACTACAGAAGTACATGGTATGGTCAGACGAAATGGTGAAGACGGGAGAGGCCCA 540

Qy         606 GATTCATGCCCACCTTGTCCGGCCCTATGTGGGCATTCATCTGCGCATTGGCTCTGACTG 665
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         541 GATTCATGCCCACCTTGTCCGGCCCTATGTGGGCATTCATCTGCGCATTGGCTCTGACTG 600

Qy         666 GAAGAACGCCTGTGCCATGCTGAAGGACGGGACTGCAGGCTCGCACTTCATGGCCTCTCC 725
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         601 GAAGAACGCCTGTGCCATGCTGAAGGACGGGACTGCAGGCTCGCACTTCATGGCCTCTCC 660

Qy         726 GCAGTGTGTGGGCTACAGCCGCAGCACAGCGGCCCCCCTCACGATGACTATGTGCCTGCC 785
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         661 GCAGTGTGTGGGCTACAGCCGCAGCACAGCGGCCCCCCTCACGATGACTATGTGCCTGCC 720

Qy         786 TGACCTGAAGGAGATCCAGAGGGCTGTGAAGCTCTGGGTGAGGTCGCTGGATGCCCAGTC 845
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         721 TGACCTGAAGGAGATCCAGAGGGCTGTGAAGCTCTGGGTGAGGTCGCTGGATGCCCAGTC 780

Qy         846 GGTCTACGTTGCTACTGATTCCGAGAGTTATGTGCCTGAGCTCCAACAGCTCTTCAAAGG 905
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         781 GGTCTACGTTGCTACTGATTCCGAGAGTTATGTGCCTGAGCTCCAACAGCTCTTCAAAGG 840

Qy         906 GAAGGTGAAGGTGGTGAGCCTGAAGCCTGAGGTGGCCCAGGTCGACCTGTACATCCTCGG 965
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         841 GAAGGTGAAGGTGGTGAGCCTGAAGCCTGAGGTGGCCCAGGTCGACCTGTACATCCTCGG 900

Qy         966 CCAAGCCGACCACTTTATTGGCAACTGTGTCTCCTCCTTCACTGCCTTTGTGAAGCGGGA 1025
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         901 CCAAGCCGACCACTTTATTGGCAACTGTGTCTCCTCCTTCACTGCCTTTGTGAAGCGGGA 960

Qy        1026 GCGGGACCTCCAGGGGAGGCCGTCTTCTTTCTTCGGCATGGACAGGCCCCCTAAGCTGCG 1085
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         961 GCGGGACCTCCAGGGGAGGCCGTCTTCTTTCTTCGGCATGGACAGGCCCCCTAAGCTGCG 1020

Qy        1086 GGACGAGTTCTGATT 1100
               |||||||||||||||
Db        1021 GGACGAGTTCTGATT 1035


RESULT 8
US-10-956-157-9986
; Sequence 9986, Application US/10956157
; Publication No. US20050118625A1
; GENERAL INFORMATION:
;  APPLICANT: Wyeth
;  APPLICANT:  Mounts, William
;  TITLE OF INVENTION: NUCLEIC ACID ARRAYS FOR DETECTING GENE EXPRESSION ASSOCIATED WIT
;  TITLE OF INVENTION:  HUMAN OSTEOARTHRITIS AND HUMAN PROTEASES
;  FILE REFERENCE: 031896-043000 (AM 101081)
;  CURRENT APPLICATION NUMBER: US/10/956,157
;  CURRENT FILING DATE:  2004-10-04
;  NUMBER OF SEQ ID NOS: 319805
;  SOFTWARE: PatentIn version 3.2
; SEQ ID NO 9986
;   LENGTH: 1400
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-10-956-157-9986

  Query Match             42.9%;  Score 471.4;  DB 22;  Length 1400;
  Best Local Similarity   99.8%;  Pred. No. 1.9e-137;
  Matches  472;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy           1 ATGCCCGCGGGCTCCTGGGACCCGGCCGGTTACCTGCTCTACTGCCCCTGCATGGGGCGC 60
               ||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          13 ATGCCTGCGGGCTCCTGGGACCCGGCCGGTTACCTGCTCTACTGCCCCTGCATGGGGCGC 72

Qy          61 TTTGGGAACCAGGCCGATCACTTCTTGGGCTCTCTGGCATTTGCAAAGCTGCTAAACCGT 120
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          73 TTTGGGAACCAGGCCGATCACTTCTTGGGCTCTCTGGCATTTGCAAAGCTGCTAAACCGT 132

Qy         121 ACCTTGGCTGTCCCTCCTTGGATTGAGTACCAGCATCACAAGCCTCCTTTCACCAACCTC 180
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         133 ACCTTGGCTGTCCCTCCTTGGATTGAGTACCAGCATCACAAGCCTCCTTTCACCAACCTC 192

Qy         181 CATGTGTCCTACCAGAAGTACTTCAAGCTGGAGCCCCTCCAGGCTTACCATCGGGTCATC 240
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         193 CATGTGTCCTACCAGAAGTACTTCAAGCTGGAGCCCCTCCAGGCTTACCATCGGGTCATC 252

Qy         241 AGCTTGGAGGATTTCATGGAGAAGCTGGCACCCACCCACTGGCCCCCTGAGAAGCGGGTG 300
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         253 AGCTTGGAGGATTTCATGGAGAAGCTGGCACCCACCCACTGGCCCCCTGAGAAGCGGGTG 312

Qy         301 GCATACTGCTTTGAGGTGGCAGCCCAGCGAAGCCCAGATAAGAAGACGTGCCCCATGAAG 360
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         313 GCATACTGCTTTGAGGTGGCAGCCCAGCGAAGCCCAGATAAGAAGACGTGCCCCATGAAG 372

Qy         361 GAAGGAAACCCCTTTGGCCCATTCTGGGATCAGTTTCATGTGAGTTTCAACAAGTCGGAG 420
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         373 GAAGGAAACCCCTTTGGCCCATTCTGGGATCAGTTTCATGTGAGTTTCAACAAGTCGGAG 432

Qy         421 CTTTTTACAGGCATTTCCTTCAGTGCTTCCTACAGAGAACAATGGAGCCAGAG 473
               |||||||||||||||||||||||||||||||||||||||||||||||||||
Db         433 CTTTTTACAGGCATTTCCTTCAGTGCTTCCTACAGAGAACAATGGAGCCAGAG 485


RESULT 9
US-10-956-157-4751
; Sequence 4751, Application US/10956157
```

```
; Publication No. US20050118625A1
; GENERAL INFORMATION:
;  APPLICANT: Wyeth
;  APPLICANT:  Mounts, William
;  TITLE OF INVENTION: NUCLEIC ACID ARRAYS FOR DETECTING GENE EXPRESSION ASSOCIATED WITH
;  TITLE OF INVENTION:  HUMAN OSTEOARTHRITIS AND HUMAN PROTEASES
;  FILE REFERENCE: 031896-043000 (AM 101081)
;  CURRENT APPLICATION NUMBER: US/10/956,157
;  CURRENT FILING DATE:  2004-10-04
;  NUMBER OF SEQ ID NOS: 319805
;  SOFTWARE: PatentIn version 3.2
; SEQ ID NO 4751
;   LENGTH: 1506
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-10-956-157-4751

  Query Match            42.9%;  Score 471.4;  DB 22;  Length 1506;
  Best Local Similarity   99.8%;  Pred. No. 1.9e-137;
  Matches  472;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 ATGCCCGCGGGCTCCTGGGACCCGGCCGGTTACCTGCTCTACTGCCCCTGCATGGGGCGC 60
              ||||| |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        119 ATGCCTGCGGGCTCCTGGGACCCGGCCGGTTACCTGCTCTACTGCCCCTGCATGGGGCGC 178

Qy         61 TTTGGGAACCAGGCCGATCACTTCTTGGGCTCTCTGGCATTTGCAAAGCTGCTAAACCGT 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        179 TTTGGGAACCAGGCCGATCACTTCTTGGGCTCTCTGGCATTTGCAAAGCTGCTAAACCGT 238

Qy        121 ACCTTGGCTGTCCCTCCTTGGATTGAGTACCAGCATCACAAGCCTCCTTTCACCAACCTC 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        239 ACCTTGGCTGTCCCTCCTTGGATTGAGTACCAGCATCACAAGCCTCCTTTCACCAACCTC 298

Qy        181 CATGTGTCCTACCAGAAGTACTTCAAGCTGGAGCCCCTCCAGGCTTACCATCGGGTCATC 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        299 CATGTGTCCTACCAGAAGTACTTCAAGCTGGAGCCCCTCCAGGCTTACCATCGGGTCATC 358

Qy        241 AGCTTGGAGGATTTCATGGAGAAGCTGGCACCCACCCACTGGCCCCCTGAGAAGCGGGTG 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        359 AGCTTGGAGGATTTCATGGAGAAGCTGGCACCCACCCACTGGCCCCCTGAGAAGCGGGTG 418

Qy        301 GCATACTGCTTTGAGGTGGCAGCCCAGCGAAGCCCAGATAAGAAGACGTGCCCCATGAAG 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        419 GCATACTGCTTTGAGGTGGCAGCCCAGCGAAGCCCAGATAAGAAGACGTGCCCCATGAAG 478

Qy        361 GAAGGAAACCCCTTTGGCCCATTCTGGGATCAGTTTCATGTGAGTTTCAACAAGTCGGAG 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        479 GAAGGAAACCCCTTTGGCCCATTCTGGGATCAGTTTCATGTGAGTTTCAACAAGTCGGAG 538

Qy        421 CTTTTTACAGGCATTTCCTTCAGTGCTTCCTACAGAGAACAATGGAGCCAGAG 473
              |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        539 CTTTTTACAGGCATTTCCTTCAGTGCTTCCTACAGAGAACAATGGAGCCAGAG 591


RESULT 10
US-09-764-891-6032
; Sequence 6032, Application US/09764891
; Publication No. US20030077808A1
; GENERAL INFORMATION:
;  APPLICANT: Rosen et al.
;  TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
;  FILE REFERENCE: PC006
;  CURRENT APPLICATION NUMBER: US/09/764,891
;  CURRENT FILING DATE:  2001-01-17
;  Prior application data removed - consult PALM or file wrapper
;  NUMBER OF SEQ ID NOS: 10231
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 6032
;   LENGTH: 10331
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-09-764-891-6032

  Query Match            22.8%;  Score 251.2;  DB 10;  Length 10331;
  Best Local Similarity   97.0%;  Pred. No. 9.4e-68;
  Matches  256;  Conservative    0;  Mismatches    8;  Indels    0;  Gaps    0;

Qy        656 GCTCTGACTGGAAGAACGCCTGTGCCATGCTGAAGGACGGGACTGCAGGCTCGCACTTCA 715
              |||||  | ||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2480 GCTCTCCGTAGAAGAACGCCTGTGCCATGCTGAAGGACGGGACTGCAGGCTCGCACTTCA 2539

Qy        716 TGGCCTCTCCGCAGTGTGTGGGCTACAGCCGCAGCACAGCGGCCCCCCTCACGATGACTA 775
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2540 TGGCCTCTCCGCAGTGTGTGGGCTACAGCCGCAGCACAGCGGCCCCCCTCACGATGACTA 2599

Qy        776 TGTGCCTGCCTGACCTGAAGGAGATCCAGAGGGCTGTGAAGCTCTGGGTGAGGTCGCTGG 835
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2600 TGTGCCTGCCTGACCTGAAGGAGATCCAGAGGGCTGTGAAGCTCTGGGTGAGGTCGCTGG 2659

Qy        836 ATGCCCAGTCGGTCTACGTTGCTACTGATTCCGAGAGTTATGTGCCTGAGCTCCAACAGC 895
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2660 ATGCCCAGTCGGTCTACGTTGCTACTGATTCCGAGAGTTATGTGCCTGAGCTCCAACAGC 2719

Qy        896 TCTTCAAAGGGAAGGTGAAGGTGG 919
              |||||||||||||||| ||||
Db       2720 TCTTCAAAGGGAAGGTATGTGTGG 2743


RESULT 11
US-09-918-995-19437
; Sequence 19437, Application US/09918995
; Publication No. US20030073623A1
; GENERAL INFORMATION:
;  APPLICANT: Hyseq, Inc.
;  TITLE OF INVENTION: NOVEL NUCLEIC ACID SEQUENCES OBTAINED
;  TITLE OF INVENTION:  FROM VARIOUS cDNA LIBRARIES
;  FILE REFERENCE: 20411-756
;  CURRENT APPLICATION NUMBER: US/09/918,995
;  CURRENT FILING DATE:  2001-07-30
;  PRIOR APPLICATION NUMBER: US/09/235,076
;  PRIOR FILING DATE: 1999-01-20
;  NUMBER OF SEQ ID NOS: 38054
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 19437
;   LENGTH: 479
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-09-918-995-19437

  Query Match            21.2%;  Score 233;  DB 10;  Length 479;
  Best Local Similarity   100.0%;  Pred. No. 1.7e-62;
  Matches  233;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        868 GAGAGTTATGTGCCTGAGCTCCAACAGCTCTTCAAAGGGAAGGTGAAGGTGGTGAGCCTG 927
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         31 GAGAGTTATGTGCCTGAGCTCCAACAGCTCTTCAAAGGGAAGGTGAAGGTGGTGAGCCTG 90

Qy        928 AAGCCTGAGGTGGCCCAGGTCGACCTGTACATCCTCGGCCAAGCCGACCACTTTATTGGC 987
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         91 AAGCCTGAGGTGGCCCAGGTCGACCTGTACATCCTCGGCCAAGCCGACCACTTTATTGGC 150

Qy        988 AACTGTGTCTCCTCCTTCACTGCCTTTGTGAAGCGGGAGCGGGACCTCCAGGGGAGGCCG 1047
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        151 AACTGTGTCTCCTCCTTCACTGCCTTTGTGAAGCGGGAGCGGGACCTCCAGGGGAGGCCG 210

Qy       1048 TCTTCTTTCTTCGGCATGGACAGGCCCCCTAAGCTGCGGGACGAGTTCTGATT 1100
              ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        211 TCTTCTTTCTTCGGCATGGACAGGCCCCCTAAGCTGCGGGACGAGTTCTGATT 263


RESULT 12
US-09-918-995-22744
; Sequence 22744, Application US/09918995
; Publication No. US20030073623A1
; GENERAL INFORMATION:
;  APPLICANT: Hyseq, Inc.
```

```
;  TITLE OF INVENTION: NOVEL NUCLEIC ACID SEQUENCES OBTAINED
;  TITLE OF INVENTION:  FROM VARIOUS cDNA LIBRARIES
;  FILE REFERENCE: 20411-756
;  CURRENT APPLICATION NUMBER: US/09/918,995
;  CURRENT FILING DATE:  2001-07-30
;  PRIOR APPLICATION NUMBER: US/09/235,076
;  PRIOR FILING DATE: 1999-01-20
;  NUMBER OF SEQ ID NOS: 38054
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 22744
;   LENGTH: 477
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(477)
;   OTHER INFORMATION: n = A,T,C or G
US-09-918-995-22744

  Query Match             20.2%;  Score 221.8;  DB 10;  Length 477;
  Best Local Similarity   96.3%;  Pred. No. 5.9e-59;
  Matches  237;  Conservative    0;  Mismatches    8;  Indels    1;  Gaps    1;

Qy        855 TGCTACTGATTCCGAGAGTTATGTGCCTGAGCTCCAACAGCTCTTCAAAGGGAAGGTGAA 914
              || | ||||   | |||| ||||||||||||||| |||| ||||||||||||||||||
Db         11 TGATGCTGAGCNCTAGAGATATGTGCCTGAGCTCC-ACAGTTCTTCAAAGGGAAGGTGAA 69

Qy        915 GGTGGTGAGCCTGAAGCCTGAGGTGGCCCAGGTCGACCTGTACATCCTCGGCCAAGCCGA 974
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         70 GGTGGTGAGCCTGAAGCCTGAGGTGGCCCAGGTCGACCTGTACATCCTCGGCCAAGCCGA 129

Qy        975 CCACTTTATTGGCAACTGTGTCTCCTCCTTCACTGCCTTTGTGAAGCGGGAGCGGGACCT 1034
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        130 CCACTTTATTGGCAACTGTGTCTCCTCCTTCACTGCCTTTGTGAAGCGGGAGCGGGACCT 189

Qy       1035 CCAGGGGAGGCCGTCTTCTTTCTTCGGCATGGACAGGCCCCCTAAGCTGCGGGACGAGTT 1094
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        190 CCAGGGGAGGCCGTCTTCTTTCTTCGGCATGGACAGGCCCCCTAAGCTGCGGGACGAGTT 249

Qy       1095 CTGATT 1100
              ||||||
Db        250 CTGATT 255


RESULT 13
US-11-097-143-18161
; Sequence 18161, Application US/11097143
; Publication No. US20050208558A1
; GENERAL INFORMATION:
;  APPLICANT: Venter, J. Craig
;  APPLICANT:  et al.
;  TITLE OF INVENTION: DETECTION KIT, SUCH AS NUCLEIC ACID
;  TITLE OF INVENTION:  ARRAYS, FOR DETECTING EXPRESSION OF 10,000 OR MORE
;  TITLE OF INVENTION:  DROSOPHILA GENES.
;  FILE REFERENCE: CL000728
;  CURRENT APPLICATION NUMBER: US/11/097,143
;  CURRENT FILING DATE:  2005-04-04
;  PRIOR APPLICATION NUMBER: 60/157,832
;  PRIOR FILING DATE: 1999-10-05
;  PRIOR APPLICATION NUMBER: 60/160,191
;  PRIOR FILING DATE: 1999-10-19
;  PRIOR APPLICATION NUMBER: 60/161,932
;  PRIOR FILING DATE: 1999-10-28
;  PRIOR APPLICATION NUMBER: 60/164,769
;  PRIOR FILING DATE: 1999-11-12
;  PRIOR APPLICATION NUMBER: 60/173,383
;  PRIOR FILING DATE: 1999-12-28
;  PRIOR APPLICATION NUMBER: 60/175,693
;  PRIOR FILING DATE: 2000-01-12
;  PRIOR APPLICATION NUMBER: 60/184,831
;  PRIOR FILING DATE: 2000-02-24
;  PRIOR APPLICATION NUMBER: 60/191,637
;  PRIOR FILING DATE: 2000-03-23
;  NUMBER OF SEQ ID NOS: 43008
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 18161
;   LENGTH: 1209
;   TYPE: DNA
;   ORGANISM: DROSOPHILA
US-11-097-143-18161

  Query Match             15.1%;  Score 166.4;  DB 26;  Length 1209;
  Best Local Similarity   51.5%;  Pred. No. 2.4e-41;
  Matches  563;  Conservative    0;  Mismatches  486;  Indels   45;  Gaps    6;

Qy         17 GGGACCCGGCCGGTTACCTGCTCTACTGCCCCTGCATGGGGCGCTTTGGGAACCAGGCCG 76
              | || ||    || |||||   |||||| || || ||||| |||||||| ||||||||||
Db         80 GCGATCCCAATGGCTACCTCACCTACTGTCCGTGTATGGGACGCTTTGGCAACCAGGCCG 139

Qy         77 ATCACTTCTTGGGCTCTCTGGCATTTGCAAAGCTGCTAAACCGTACCTTGGCTGTCCCTC 136
              | |||||| |||| ||  |||| || || |||  ||| || || ||| ||    | || |
Db        140 ACCACTTCCTGGGATCATTGGCCTTCGCCAAGGCGCTTAATCGCACCCTGATCCTGCCGC 199

Qy        137 CTTGGATTGAGTACCAGCATCACAAGCCTCCTTTCACCAACCTCCATGTGTCCTACCAGA 196
              | ||| | ||||      ||| | | |     |      ||  || || | | | | |
Db        200 CGTGGGTGGAGT------ATCGTAGGGGTGAACTGCGATCCCGTCAGGTACCGTTCAACA 253

Qy        197 AGTACTTCAAGCTGGAGCCCCTCCAGGCTTACCATCGGGTCATCAGCTTGGAGGATTTCA 256
                || ||  | |||||||||| ||| |||||||| ||||||| | ||| |||||||
Db        254 CATATTTTGAAGTGGAGCCCCTGAAGGAATACCATCGCGTCATCACCATGGCAGATTTCA 313

Qy        257 TGGAGAAGCTGGCACCCACCCACTGGCCCCCTGAGAAGCGGGTGGCATACTGCTTTGAGG 316
              ||  | | |||||   |  |  |||||       | |||| ||| |||  ||||   |||
Db        314 TGTGGCACCTGGCCGACGACATTTGGCCAGAATCGGAGCGAGTGTCATTTTGCTACAAGG 373

Qy        317 --------------TGGCAGCCCAGCGAAGCCCAGATAAGAAGACGTGCCCCATGAAGG 361
                           |   |   ||      ||||| |||   |  |||| |  ||||
Db        374 AACGATATAGCCTTCAGCAGGAGAAGAACGATCCAGACAAGCCCAATTGCCACGCCAAGG 433

Qy        362 AAGGAAACCCCTTTGGCCCATTCTGGGATCAGTTTCATGTGAGTTTCAACAAGTCGGAGC 421
              | || || || ||||| || || |||||     |||||  |    ||       ||| |||
Db        434 ATGGCAATCCTTTTGGTCCCTTTTGGGACACTTTTCACATAGACTTTGTGCGGTCAGAGT 493

Qy        422 TTTTTACAGGCATTTCCTTCAGTGCTTCCTACAG---------AGAACAATGGAGCCAGA 472
              | | | |    ||   ||   ||     | ||         |   | |||    |
Db        494 TCTATGCGCCACTTCATTTTGATGTGCATCATAGCAACGAGGCTGCCAAGTGGCAGACCA 553

Qy        473 GATTTTCTCCAAAGGAACATCCGGTGCTTGCCCTGCCAGGAGCCCCAGCCCAGTTCCCCG 532
              | | || || |   | |||| || || ||  |  | ||||| || ||    || || |
Db        554 AATATCCTGCAGAATCATATCCCGTACTCGCGTTCACCGGAGCTCCGGCTAGTTTTCCTG 613

Qy        533 TCCTAGAAGAACACAGGCCACTACAGAAGTACATGGTATGGTCAGACGAAATGGTGAAGA 592
              | |    ||| || |    || |||   ||| ||  |||  |           | |
Db        614 TTCAGCTAGAGAACTGCAAGCTGCAGCGCTACTTGCAGTGGAGTCAACGGTATAGGGAAG 673

Qy        593 CGGGAGAGGCCCAGATTCATGCCCACCTTGTCCGGC---CCTATGTGGGCATTCATCTGC 649
              |     |||     || |  |  ||  |     |||    ||| | ||||||||||||||
Db        674 CATCTAAGGATTTCATCCGAGAGCAGTTGCCTCGGGGTGCCTTTTTGGGCATTCATCTGC 733

Qy        650 GCATTGGCTCTGACTGGAAGAACGCCTGTGCCATGCTGAAGGACGG---------GACTG 700
              |||  ||    || |||  ||  |||||||       | |||||  |         |  ||
Db        734 GCAACGGTATCGATTGGGTGAGAGCCTGTGAGCACGTCAAGGATAGCCAGCATCTGTTTG 793

Qy        701 CAGGCTCGCACTTCATGGCCTCTCCGCAGTGTGTGGGCTACAGCCGCAGCACAGCGGCCC 760
              |     |||| | | ||| || |     |          ||     | |          |
Db        794 CCTCGCCGCAGTGCTTGGGCTATAAAAATGAACGTGGTGCACTCTACCCGGAGCTCTGCA 853

Qy        761 CCCTCACGATGACTATGTGCCTGCCTGACCTGAAGGAGATCCAGAGGGCTGTGAAGCTCT 820
              | | | | |     |  | | |   | || |||   | |   | |  |||    |
Db        854 TGCCCTCCAAGGAGGCGATCATCCGCCAGCTAAAGAGAACCATTAAGAACGTGCGCCAAA 913

Qy        821 GGGTGAGGTCGCTGGATGCCCAGTCGGTCTACGTTGCTACTGATTCCGAGAGTTATGTGC 880
              |  |       ||   | | || || | ||| ||  | || ||| |         |
Db        914 CTCAGCCGGACAACGAAATCAAATCAGTTTTCGTGGCGTCAGACTCCAATCACATGATTG 973

Qy        881 CTGAGCTCCAACAGCTCTTCAAAGGGAAGGTGAAGGTGGTGAGCCTGAAGCCTGAGGTGG 940
```

```
Db        974 GTGAACTAAACACGGCCCTTAGTCGCATGGGCATCAGTGTGCACAAGCTGCCGGAGGATG 1033

Qy        941 ---CCCAGGTCGACCTGTACATCCTCGGCCAAGCCGACCACTTTATTGGCAACTGTGTCT 997

Db       1034 ATCCTTACCTGGACTTGGCCATTCTCGGACAGTCGAACCACTTTATCGGCAACTGTATAT 1093

Qy        998 CCTCCTTCACTGCCTTTGTGAAGCGGGAGCGGGACCTCCAGGGGAGGCCGTCTTCTTTCT 1057

Db       1094 CCTCTTACTCGGCATTCGAAAAAAGGGAACGAGATGTGCACGGTTTTCCATCGTACTTCT 1153

Qy       1058 TCGGCATGGACAGG 1071

Db       1154 GGGGATTCCCCAAG 1167


RESULT 14
US-11-097-143-18160/c
; Sequence 18160, Application US/11097143
; Publication No. US20050208558A1
; GENERAL INFORMATION:
;  APPLICANT: Venter, J. Craig
;  APPLICANT:  et al.
;  TITLE OF INVENTION: DETECTION KIT, SUCH AS NUCLEIC ACID
;  TITLE OF INVENTION:  ARRAYS, FOR DETECTING EXPRESSION OF 10,000 OR MORE
;  TITLE OF INVENTION:  DROSOPHILA GENES.
;  FILE REFERENCE: CL000728
;  CURRENT APPLICATION NUMBER: US/11/097,143
;  CURRENT FILING DATE:  2005-04-04
;  PRIOR APPLICATION NUMBER: 60/157,832
;  PRIOR FILING DATE: 1999-10-05
;  PRIOR APPLICATION NUMBER: 60/160,191
;  PRIOR FILING DATE: 1999-10-19
;  PRIOR APPLICATION NUMBER: 60/161,932
;  PRIOR FILING DATE: 1999-10-28
;  PRIOR APPLICATION NUMBER: 60/164,769
;  PRIOR FILING DATE: 1999-11-12
;  PRIOR APPLICATION NUMBER: 60/173,383
;  PRIOR FILING DATE: 1999-12-28
;  PRIOR APPLICATION NUMBER: 60/175,693
;  PRIOR FILING DATE: 2000-01-12
;  PRIOR APPLICATION NUMBER: 60/184,831
;  PRIOR FILING DATE: 2000-02-24
;  PRIOR APPLICATION NUMBER: 60/191,637
;  PRIOR FILING DATE: 2000-03-23
;  NUMBER OF SEQ ID NOS: 43008
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 18160
;   LENGTH: 3264
;   TYPE: DNA
;   ORGANISM: DROSOPHILA
US-11-097-143-18160

  Query Match            13.7%;  Score 150.4;  DB 26;  Length 3264;
  Best Local Similarity   50.9%;  Pred. No. 3.7e-36;
  Matches  541;  Conservative    0;  Mismatches  476;  Indels   45;  Gaps    6;

Qy         49 TGCATGGGGCGCTTTGGGAACCAGGCCGATCACTTCTTGGGCTCTCTGGCATTTGCAAAG 108

Db       2098 TTCAAAGGACGCTTTGGCAACCAGGCCGACCACTTCCTGGGATCATTGGCCTTCGCCAAG 2039

Qy        109 CTGCTAAACCGTACCTTGGCTGTCCCTCCTTGGATTGAGTACCAGCATCACAAGCCTCCT 168

Db       2038 GCGCTTAATCGCACCCTGATCCTGCCGCCGTGGGTGGAGT------ATCGTAGGGGTGAA 1985

Qy        169 TTCACCAACCTCCATGTGTCCTACCAGAAGTACTTCAAGCTGGAGCCCCTCCAGGCTTAC 228

Db       1984 CTGCGATCCCGTCAGGTACCGTTCAACACATATTTTGAAGTGGAGCCCCTGAAGGAATAC 1925

Qy        229 CATCGGGTCATCAGCTTGGAGGATTTCATGGAGAAGCTGGCACCCACCCACTGGCCCCCT 288

Db       1924 CATCGCGTCATCACCATGGCAGATTTCATGTGGCACCTGGCCGACGACATTTGGCCAGAA 1865

Qy        289 GAGAAGCGGGTGGCATACTGCTTTGAGG----------TGGCAGCCCAGCGAAGC 333

Db       1864 TCGGAGCGAGTGTCATTTTGCTACAAGGAACGATATAGCCTTCAGCAGGAGAAGAACGAT 1805

Qy        334 CCAGATAAGAAGACGTGCCCCATGAAGGAAGGAAACCCCTTTGGCCCATTCTGGGATCAG 393

Db       1804 CCAGACAAGCCCAATTGCCACGCCAAGGATGGCAATCCTTTTGGTCCCTTTTGGGACACT 1745

Qy        394 TTTCATGTGAGTTTCAACAAGTCGGAGCTTTTTACAGGCATTTCCTTCAGTGCTTCCTAC 453

Db       1744 TTTCACATAGACTTTGTGCGGTCAGAGTTCTATGCGCCACTTCATTTTGATGTGCATCAT 1685

Qy        454 AG---------AGAACAATGGAGCCAGAGATTTTCTCCAAAGGAACATCCGGTGCTTGCC 504

Db       1684 AGCAACGAGGCTGCCAAGTGGCAGACCAAATATCCTGCAGAATCATATCCCGTACTCGCG 1625

Qy        505 CTGCCAGGAGCCCCAGCCCAGTTCCCCGTCCTAGAAGAACACAGGCCACTACAGAAGTAC 564

Db       1624 TTCACCGGAGCTCCGGCTAGTTTTCCTGTTCAGCTAGAGAACTGCAAGCTGCAGCGCTAC 1565

Qy        565 ATGGTATGGTCAGACGAAATGGTGAAGACGGGAGAGGCCCAGATTCATGCCCACCTTGTC 624

Db       1564 TTGCAGTGGAGTCAACGGTATAGGGAAGCATCTAAGGATTTCATCCGAGAGCAGTTGCCT 1505

Qy        625 CGG---CCCTATGTGGGCATTCATCTGCGCATTGGCTCTGACTGGAAGAACGCCTGTGCC 681

Db       1504 CGGGGTGCCTTTTTGGGCATTCATCTGCGCAACGGTATCGATTGGGTGAGAGCCTGTGAG 1445

Qy        682 ATGCTGAAGGACGG---------GACTGCAGGCTCGCACTTCATGGCCTCTCCGCAGTGT 732

Db       1444 CACGTCAAGGATAGCCAGCATCTGTTTGCCTCGCCGCAGTGCTTGGGCTATAAAAATGAA 1385

Qy        733 GTGGGCTACAGCCGCAGCACAGCGGCCCCCCTCACGATGACTATGTGCCTGCCTGACCTG 792

Db       1384 CGTGGTGCACTCTACCCGGAGCTCTGCATGCCCTCCAAGGAGGCGATCATCCGCCAGCTA 1325

Qy        793 AAGGAGATCCAGAGGGCTGTGAAGCTCTGGGTGAGGTCGCTGGATGCCCAGTCGGTCTAC 852

Db       1324 AAGAGAACCATTAAGAACGTGCGCCAAACTCAGCCGGACAACGAAATCAAATCAGTTTTC 1265

Qy        853 GTTGCTACTGATTCCGAGAGTTATGTGCCTGAGCTCCAACAGCTCTTCAAAGGGAAGGTG 912

Db       1264 GTGGCGTCAGACTCCAATCACATGATTGGTGAACTAAACACGGCCCTTAGTCGCATGGGC 1205

Qy        913 AAGGTGGTGAGCCTGAAGCCTGAGGTGG---CCCAGGTCGACCTGTACATCCTCGGCCAA 969

Db       1204 ATCAGTGTGCACAAGCTGCCGGAGGATGATCCTTACCTGGACTTGGCCATTCTCGGACAG 1145

Qy        970 GCCGACCACTTTATTGGCAACTGTGTCTCCTCCTTCACTGCCTTTGTGAAGCGGGAGCGG 1029

Db       1144 TCGAACCACTTTATCGGCAACTGTATATCCTCTTACTCGGCATTCGAAAAAAGGGAACGA 1085

Qy       1030 GACCTCCAGGGGAGGCCGTCTTCTTTCTTCCGGCATGGACAGG 1071

Db       1084 GATGTGCACGGTTTTCCATCGTACTTCTGGGGATTCCCCAAG 1043


RESULT 15
US-09-783-590-11501
; Sequence 11501, Application US/09783590
; Patent No. US20020110850A1
; GENERAL INFORMATION:
;  APPLICANT: Dillon, Patrick J.
;  APPLICANT:  Haseltine, William A.
;  APPLICANT:  Li, Haodong
;  APPLICANT:  Rosen, Craig A.
;  APPLICANT:  Ruben, Steven M.
;  TITLE OF INVENTION: Human Genes, Sequences, and Expression Products 16.2
;  FILE REFERENCE: PO-16.2C1
;  CURRENT APPLICATION NUMBER: US/09/783,590
;  CURRENT FILING DATE:  2000-02-15
;  PRIOR APPLICATION NUMBER: 08/420,856
;  PRIOR FILING DATE: 1995-04-12
```

; PRIOR APPLICATION NUMBER: 08/346,731
; PRIOR FILING DATE: 1994-11-21
; NUMBER OF SEQ ID NOS: 12485
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 11501
; LENGTH: 402
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (9)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (19)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (22)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (46)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (272)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (301)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (320)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (333)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (342)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (374)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (386)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (390)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (400)
; OTHER INFORMATION: n equals a,t,g, or c
US-09-783-590-11501

Query Match 10.4%; Score 114.6; DB 9; Length 402;
Best Local Similarity 83.9%; Pred. No. 3.4e-25;
Matches 177; Conservative 0; Mismatches 23; Indels 11; Gaps 4;

```
Qy        178 CTCCATGTGTCCTACCAGAAGTACTTCAAGCTGGAGCCCCTCCAGGCTTACCATCGGGTC 237
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        161 CTCCATGTGTCCTACCAGAAGTACTTCAAGCTGGAGCCCCTCCAGGCTTACCATCGGGTC 220

Qy        238 ATCAGCTT-GGAGGATTTCATGGAGAAGCTGGCACCCACCCACTGGCCCCCT---GAGAA 293
              |||||||| |||||||||||||||||||||||||||||||||||| ||||    ||
Db        221 ATCAGCTTGGGAGGATTTCATGGAGAAGCTGGCACCCACCCACTGGGCCCCNTGAGAAGC 280

Qy        294 GCGGGTGGCATACTGCTTTGA--GGTGGCAGCCCAGCGAAGCCCAGAT-----AAGAAGA 346
              | ||  |||||| ||||||     |  |||||||||||| ||||||||     ||| |
Db        281 GGGGTGGGCATAATGCTTTTNAGGTGGGCAGCCCAGCGANGCCCAGATTAAGNAAGGACG 340

Qy        347 CGTGCCCCATGAAGGAAGGAAACCCCTTTGG 377
                  |||||||||||||||| |||||| |
Db        341 TNGACCCCATGAAGGAAGGAAAACCCTTTTG 371
```

Search completed: November 9, 2005, 15:51:17
Job time : 988.698 secs

GenCore version 5.1.6

OM nucleic - nucleic search, using sw model

Run on: November 9, 2005, 02:15:54 ; Search time 199.044 Seconds
(without alignments)
9042.764 Million cell updates/sec

Title: US-09-774-954-16
Perfect score: 1100
Sequence: 1 ATGCCCGCGGGCTCCTGGGA.....CTGCGGGACGAGTTCTGATT 1100

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1202784 seqs, 818138359 residues

Total number of hits satisfying chosen parameters: 2405568

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_NA:*
1: /cgn2_6/ptodata/1/ina/5A_COMB.seq:*
2: /cgn2_6/ptodata/1/ina/5B_COMB.seq:*
3: /cgn2_6/ptodata/1/ina/6A_COMB.seq:*
4: /cgn2_6/ptodata/1/ina/6B_COMB.seq:*
5: /cgn2_6/ptodata/1/ina/PCTUS_COMB.seq:*
6: /cgn2_6/ptodata/1/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	1100	100.0	1100	3	US-08-978-741-16	Sequence 16, Appl
	2	1100	100.0	1100	3	US-09-333-729A-16	Sequence 16, Appl
	3	1100	100.0	1514	3	US-08-978-741-1	Sequence 1, Appli
	4	1100	100.0	1514	3	US-09-333-729A-2	Sequence 2, Appli
	5	1098.4	99.9	1300	3	US-08-978-741-4	Sequence 4, Appli
	6	1098.4	99.9	1300	3	US-09-333-729A-6	Sequence 6, Appli
	7	1098.4	99.9	11284	3	US-08-978-741-5	Sequence 5, Appli
	8	1033.4	93.9	5009	3	US-08-978-741-7	Sequence 7, Appli
	9	1033.4	93.9	5009	3	US-09-333-729A-8	Sequence 8, Appli
	10	168	15.3	1320	4	US-09-270-767-14353	Sequence 14353, A
	11	62.2	5.7	722	4	US-09-270-767-30507	Sequence 30507, A
c	12	58.6	5.3	998	4	US-09-270-767-28905	Sequence 28905, A
c	13	58.6	5.3	2034	4	US-09-270-767-13022	Sequence 13022, A
c	14	56	5.1	19227	4	US-09-949-016-12127	Sequence 12127, A
c	15	56	5.1	19228	4	US-09-949-016-16285	Sequence 16285, A
c	16	41.2	3.7	114793	4	US-10-148-806-3	Sequence 3, Appli
	17	40	3.6	40	3	US-08-978-741-10	Sequence 10, Appl
	18	40	3.6	40	3	US-08-978-741-12	Sequence 12, Appl
	19	40	3.6	40	3	US-09-333-729A-9	Sequence 9, Appli
	20	39	3.5	1131	4	US-09-252-991A-76	Sequence 76, Appl
	21	39	3.5	1353	4	US-09-252-991A-88	Sequence 88, Appl
	22	38.6	3.5	2358	4	US-09-919-039-190	Sequence 190, App
	23	37.8	3.4	2154	4	US-09-902-540-4557	Sequence 4557, Ap
	24	37.8	3.4	24905	4	US-09-902-540-1225	Sequence 1225, Ap
c	25	37.8	3.4	4403765	3	US-09-103-840A-2	Sequence 2, Appli
c	26	37.8	3.4	4411529	3	US-09-103-840A-1	Sequence 1, Appli
c	27	37.4	3.4	18037	4	US-09-949-016-13678	Sequence 13678, A
c	28	37.4	3.4	77536	4	US-09-410-551B-1	Sequence 1, Appli
c	29	37.4	3.4	77536	4	US-09-940-316B-1	Sequence 1, Appli
	30	37.2	3.4	933	4	US-09-252-991A-11943	Sequence 11943, A
c	31	37.2	3.4	1218	4	US-09-252-991A-12011	Sequence 12011, A
	32	37.2	3.4	1650	4	US-09-252-991A-11977	Sequence 11977, A
	33	37.2	3.4	102053	4	US-09-949-016-13025	Sequence 13025, A
	34	37	3.4	1843	1	US-07-918-023-1	Sequence 1, Appli
	35	37	3.4	5962	6	5386025-5	Patent No. 5386025
	36	37	3.4	5962	6	5386025-5	Patent No. 5386025
	37	37	3.4	5975	1	US-08-404-354B-1	Sequence 1, Appli
	38	37	3.4	5975	1	US-08-314-083B-1	Sequence 1, Appli
	39	37	3.4	5975	1	US-08-435-675B-1	Sequence 1, Appli
	40	37	3.4	5975	1	US-08-336-257A-3	Sequence 3, Appli
	41	37	3.4	5975	3	US-08-884-599-1	Sequence 1, Appli
	42	36.8	3.3	3591	4	US-10-237-551-191	Sequence 191, App
	43	36.8	3.3	3591	4	US-10-237-551-221	Sequence 221, App
	44	36.8	3.3	3981	4	US-09-294-298A-5	Sequence 5, Appli
c	45	36.8	3.3	3981	4	US-09-294-298A-24	Sequence 24, Appl

ALIGNMENTS

RESULT 1
US-08-978-741-16
; Sequence 16, Application US/08978741
; Patent No. 6100076
; GENERAL INFORMATION:
; APPLICANT: Yang Wang, Michael W. Spellman
; TITLE OF INVENTION: O-Fucosyltransferase
; NUMBER OF SEQUENCES: 17
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genentech, Inc.
; STREET: 1 DNA Way
; CITY: South San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94080
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5 inch, 1.44 Mb floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: WinPatin (Genentech)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/978,741
; FILING DATE: 26-No. 6100076-1997
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/792498
; FILING DATE: 31
; ATTORNEY/AGENT INFORMATION:
; NAME: Svoboda, Craig G.
; REGISTRATION NUMBER: 39,044
; REFERENCE/DOCKET NUMBER: P1041P1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 650/225-1489
; TELEFAX: 650/952-9881
; INFORMATION FOR SEQ ID NO: 16:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1100 base pairs
; TYPE: Nucleic Acid
; STRANDEDNESS: Single
; TOPOLOGY: Linear
US-08-978-741-16

Query Match 100.0%; Score 1100; DB 3; Length 1100;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1100; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 ATGCCCGCGGGCTCCTGGGACCCGGCCGGTTACCTGCTCTACTGCCCCTGCATGGGGCGC 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGCCCGCGGGCTCCTGGGACCCGGCCGGTTACCTGCTCTACTGCCCCTGCATGGGGCGC 60
```

```
Qy        61 TTTGGGAACCAGGCCGATCACTTCTTGGGCTCTCTGGCATTTGCAAAGCTGCTAAACCGT 120
Db        61 TTTGGGAACCAGGCCGATCACTTCTTGGGCTCTCTGGCATTTGCAAAGCTGCTAAACCGT 120

Qy       121 ACCTTGGCTGTCCCTCCTTGGATTGAGTACCAGCATCACAAGCCTCCTTTCACCAACCTC 180
Db       121 ACCTTGGCTGTCCCTCCTTGGATTGAGTACCAGCATCACAAGCCTCCTTTCACCAACCTC 180

Qy       181 CATGTGTCCTACCAGAAGTACTTCAAGCTGGAGCCCCTCCAGGCTTACCATCGGGTCATC 240
Db       181 CATGTGTCCTACCAGAAGTACTTCAAGCTGGAGCCCCTCCAGGCTTACCATCGGGTCATC 240

Qy       241 AGCTTGGAGGATTTCATGGAGAAGCTGGCACCCACCCACTGGCCCCCTGAGAAGCGGGTG 300
Db       241 AGCTTGGAGGATTTCATGGAGAAGCTGGCACCCACCCACTGGCCCCCTGAGAAGCGGGTG 300

Qy       301 GCATACTGCTTTGAGGTGGCAGCCCAGCGAAGCCCAGATAAGAAGACGTGCCCCATGAAG 360
Db       301 GCATACTGCTTTGAGGTGGCAGCCCAGCGAAGCCCAGATAAGAAGACGTGCCCCATGAAG 360

Qy       361 GAAGGAAACCCCTTTGGCCCATTCTGGGATCAGTTTCATGTGAGTTTCAACAAGTCGGAG 420
Db       361 GAAGGAAACCCCTTTGGCCCATTCTGGGATCAGTTTCATGTGAGTTTCAACAAGTCGGAG 420

Qy       421 CTTTTTACAGGCATTTCCTTCAGTGCTTCCTACAGAGAACAATGGAGCCAGAGATTTTCT 480
Db       421 CTTTTTACAGGCATTTCCTTCAGTGCTTCCTACAGAGAACAATGGAGCCAGAGATTTTCT 480

Qy       481 CCAAAGGAACATCCGGTGCTTGCCCTGCCAGGAGCCCCAGCCCAGTTCCCCGTCCTAGAA 540
Db       481 CCAAAGGAACATCCGGTGCTTGCCCTGCCAGGAGCCCCAGCCCAGTTCCCCGTCCTAGAA 540

Qy       541 GAACACAGGCCACTACAGAAGTACATGGTATGGTCAGACGAAATGGTGAAGACGGGAGAG 600
Db       541 GAACACAGGCCACTACAGAAGTACATGGTATGGTCAGACGAAATGGTGAAGACGGGAGAG 600

Qy       601 GCCCAGATTCATGCCCACCTTGTCCGGCCCTATGTGGGCATTCATCTGCGCATTGGCTCT 660
Db       601 GCCCAGATTCATGCCCACCTTGTCCGGCCCTATGTGGGCATTCATCTGCGCATTGGCTCT 660

Qy       661 GACTGGAAGAACGCCTGTGCCATGCTGAAGGACGGGACTGCAGGCTCGCACTTCATGGCC 720
Db       661 GACTGGAAGAACGCCTGTGCCATGCTGAAGGACGGGACTGCAGGCTCGCACTTCATGGCC 720

Qy       721 TCTCCGCAGTGTGTGGGCTACAGCCGCAGCACAGCGGCCCCCCTCACGATGACTATGTGC 780
Db       721 TCTCCGCAGTGTGTGGGCTACAGCCGCAGCACAGCGGCCCCCCTCACGATGACTATGTGC 780

Qy       781 CTGCCTGACCTGAAGGAGATCCAGAGGGCTGTGAAGCTCTGGGTGAGGTCGCTGGATGCC 840
Db       781 CTGCCTGACCTGAAGGAGATCCAGAGGGCTGTGAAGCTCTGGGTGAGGTCGCTGGATGCC 840

Qy       841 CAGTCGGTCTACGTTGCTACTGATTCCGAGAGTTATGTGCCTGAGCTCCAACAGCTCTTC 900
Db       841 CAGTCGGTCTACGTTGCTACTGATTCCGAGAGTTATGTGCCTGAGCTCCAACAGCTCTTC 900

Qy       901 AAAGGGAAGGTGAAGGTGGTGAGCCTGAAGCCTGAGGTGGCCCAGGTCGACCTGTACATC 960
Db       901 AAAGGGAAGGTGAAGGTGGTGAGCCTGAAGCCTGAGGTGGCCCAGGTCGACCTGTACATC 960

Qy       961 CTCGGCCAAGCCGACCACTTTATTGGCAACTGTGTCTCCTCCTTCACTGCCTTTGTGAAG 1020
Db       961 CTCGGCCAAGCCGACCACTTTATTGGCAACTGTGTCTCCTCCTTCACTGCCTTTGTGAAG 1020

Qy      1021 CGGGAGCGGGACCTCCAGGGGAGGCCGTCTTCTTTCTTCGGCATGGACAGGCCCCCTAAG 1080
Db      1021 CGGGAGCGGGACCTCCAGGGGAGGCCGTCTTCTTTCTTCGGCATGGACAGGCCCCCTAAG 1080

Qy      1081 CTGCGGGACGAGTTCTGATT 1100
Db      1081 CTGCGGGACGAGTTCTGATT 1100
```

```
RESULT 2
US-09-333-729A-16
; Sequence 16, Application US/09333729A
; Patent No. 6270987
; GENERAL INFORMATION:
;  APPLICANT:  Wang,Yang
;  APPLICANT:  Spellman,Michael W.
;  TITLE OF INVENTION:  O-Fucosyltransferase
;  FILE REFERENCE:  P1041P1D1-Substitute
;  CURRENT APPLICATION NUMBER: US/09/333,729A
;  CURRENT FILING DATE:  1999-06-15
;  PRIOR APPLICATION NUMBER:  US 08/798,741
;  PRIOR FILING DATE:  1997-11-26
;  NUMBER OF SEQ ID NOS:  21
; SEQ ID NO 16
;   LENGTH: 1100
;   TYPE: DNA
;   ORGANISM: Homo Sapien
US-09-333-729A-16

  Query Match           100.0%;  Score 1100;  DB 3;  Length 1100;
  Best Local Similarity  100.0%;  Pred. No. 0;
  Matches 1100;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 ATGCCCGCGGGGCTCCTGGGACCCGGCCGGTTACCTGCTCTACTGCCCCTGCATGGGGCGC 60
Db          1 ATGCCCGCGGGGCTCCTGGGACCCGGCCGGTTACCTGCTCTACTGCCCCTGCATGGGGCGC 60

Qy         61 TTTGGGAACCAGGCCGATCACTTCTTGGGCTCTCTGGCATTTGCAAAGCTGCTAAACCGT 120
Db         61 TTTGGGAACCAGGCCGATCACTTCTTGGGCTCTCTGGCATTTGCAAAGCTGCTAAACCGT 120

Qy        121 ACCTTGGCTGTCCCTCCTTGGATTGAGTACCAGCATCACAAGCCTCCTTTCACCAACCTC 180
Db        121 ACCTTGGCTGTCCCTCCTTGGATTGAGTACCAGCATCACAAGCCTCCTTTCACCAACCTC 180

Qy        181 CATGTGTCCTACCAGAAGTACTTCAAGCTGGAGCCCCTCCAGGCTTACCATCGGGTCATC 240
Db        181 CATGTGTCCTACCAGAAGTACTTCAAGCTGGAGCCCCTCCAGGCTTACCATCGGGTCATC 240

Qy        241 AGCTTGGAGGATTTCATGGAGAAGCTGGCACCCACCCACTGGCCCCCTGAGAAGCGGGTG 300
Db        241 AGCTTGGAGGATTTCATGGAGAAGCTGGCACCCACCCACTGGCCCCCTGAGAAGCGGGTG 300

Qy        301 GCATACTGCTTTGAGGTGGCAGCCCAGCGAAGCCCAGATAAGAAGACGTGCCCCATGAAG 360
Db        301 GCATACTGCTTTGAGGTGGCAGCCCAGCGAAGCCCAGATAAGAAGACGTGCCCCATGAAG 360

Qy        361 GAAGGAAACCCCTTTGGCCCATTCTGGGATCAGTTTCATGTGAGTTTCAACAAGTCGGAG 420
Db        361 GAAGGAAACCCCTTTGGCCCATTCTGGGATCAGTTTCATGTGAGTTTCAACAAGTCGGAG 420

Qy        421 CTTTTTACAGGCATTTCCTTCAGTGCTTCCTACAGAGAACAATGGAGCCAGAGATTTTCT 480
Db        421 CTTTTTACAGGCATTTCCTTCAGTGCTTCCTACAGAGAACAATGGAGCCAGAGATTTTCT 480

Qy        481 CCAAAGGAACATCCGGTGCTTGCCCTGCCAGGAGCCCCAGCCCAGTTCCCCGTCCTAGAA 540
Db        481 CCAAAGGAACATCCGGTGCTTGCCCTGCCAGGAGCCCCAGCCCAGTTCCCCGTCCTAGAA 540

Qy        541 GAACACAGGCCACTACAGAAGTACATGGTATGGTCAGACGAAATGGTGAAGACGGGAGAG 600
Db        541 GAACACAGGCCACTACAGAAGTACATGGTATGGTCAGACGAAATGGTGAAGACGGGAGAG 600

Qy        601 GCCCAGATTCATGCCCACCTTGTCCGGCCCTATGTGGGCATTCATCTGCGCATTGGCTCT 660
Db        601 GCCCAGATTCATGCCCACCTTGTCCGGCCCTATGTGGGCATTCATCTGCGCATTGGCTCT 660

Qy        661 GACTGGAAGAACGCCTGTGCCATGCTGAAGGACGGGACTGCAGGCTCGCACTTCATGGCC 720
Db        661 GACTGGAAGAACGCCTGTGCCATGCTGAAGGACGGGACTGCAGGCTCGCACTTCATGGCC 720

Qy        721 TCTCCGCAGTGTGTGGGCTACAGCCGCAGCACAGCGGCCCCCCTCACGATGACTATGTGC 780
```

```
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 TCTCCGCAGTGTGTGGGCTACAGCCGCAGCACAGCGGCCCCCCTCACGATGACTATGTGC 780

Qy        781 CTGCCTGACCTGAAGGAGATCCAGAGGGCTGTGAAGCTCTGGGTGAGGTCGCTGGATGCC 840
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 CTGCCTGACCTGAAGGAGATCCAGAGGGCTGTGAAGCTCTGGGTGAGGTCGCTGGATGCC 840

Qy        841 CAGTCGGTCTACGTTGCTACTGATTCCGAGAGTTATGTGCCTGAGCTCCAACAGCTCTTC 900
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 CAGTCGGTCTACGTTGCTACTGATTCCGAGAGTTATGTGCCTGAGCTCCAACAGCTCTTC 900

Qy        901 AAAGGGAAGGTGAAGGTGGTGAGCCTGAAGCCTGAGGTGGCCCAGGTCGACCTGTACATC 960
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 AAAGGGAAGGTGAAGGTGGTGAGCCTGAAGCCTGAGGTGGCCCAGGTCGACCTGTACATC 960

Qy        961 CTCGGCCAAGCCGACCACTTTATTGGCAACTGTGTCTCCTCCTTCACTGCCTTTGTGAAG 1020
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 CTCGGCCAAGCCGACCACTTTATTGGCAACTGTGTCTCCTCCTTCACTGCCTTTGTGAAG 1020

Qy       1021 CGGGAGCGGGACCTCCAGGGGAGGCCGTCTTCTTTCTTCGGCATGGACAGGCCCCCTAAG 1080
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1021 CGGGAGCGGGACCTCCAGGGGAGGCCGTCTTCTTTCTTCGGCATGGACAGGCCCCCTAAG 1080

Qy       1081 CTGCGGGACGAGTTCTGATT 1100
                ||||||||||||||||||||
Db       1081 CTGCGGGACGAGTTCTGATT 1100


RESULT 3
US-08-978-741-1
; Sequence 1, Application US/08978741
; Patent No. 6100076
;  GENERAL INFORMATION:
;    APPLICANT:  Yang Wang, Michael W. Spellman
;    TITLE OF INVENTION:  O-Fucosyltransferase
;    NUMBER OF SEQUENCES:  17
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Genentech, Inc.
;      STREET:  1 DNA Way
;      CITY:  South San Francisco
;      STATE:  California
;      COUNTRY:  USA
;      ZIP:  94080
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  3.5 inch, 1.44 Mb floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  WinPatin (Genentech)
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/978,741
;      FILING DATE:  26-No. 6100076-1997
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/792498
;      FILING DATE:  31
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Svoboda, Craig G.
;      REGISTRATION NUMBER:  39,044
;      REFERENCE/DOCKET NUMBER:  P1041P1
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  650/225-1489
;      TELEFAX:  650/952-9881
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1514 base pairs
;      TYPE:  Nucleic Acid
;      STRANDEDNESS:  Single
;      TOPOLOGY:  Linear
US-08-978-741-1

  Query Match             100.0%;  Score 1100;  DB 3;  Length 1514;
  Best Local Similarity   100.0%;  Pred. No. 0;
  Matches 1100;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy           1 ATGCCCGCGGGCTCCTGGGACCCGGCCGGTTACCTGCTCTACTGCCCCTGCATGGGGCGC 60
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db           1 ATGCCCGCGGGCTCCTGGGACCCGGCCGGTTACCTGCTCTACTGCCCCTGCATGGGGCGC 60

Qy          61 TTTGGGAACCAGGCCGATCACTTCTTGGGCTCTCTGGCATTTGCAAAGCTGCTAAACCGT 120
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          61 TTTGGGAACCAGGCCGATCACTTCTTGGGCTCTCTGGCATTTGCAAAGCTGCTAAACCGT 120

Qy         121 ACCTTGGCTGTCCCTCCTTGGATTGAGTACCAGCATCACAAGCCTCCTTTCACCAACCTC 180
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         121 ACCTTGGCTGTCCCTCCTTGGATTGAGTACCAGCATCACAAGCCTCCTTTCACCAACCTC 180

Qy         181 CATGTGTCCTACCAGAAGTACTTCAAGCTGGAGCCCCTCCAGGCTTACCATCGGGTCATC 240
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         181 CATGTGTCCTACCAGAAGTACTTCAAGCTGGAGCCCCTCCAGGCTTACCATCGGGTCATC 240

Qy         241 AGCTTGGAGGATTTCATGGAGAAGCTGGCACCCACCCACTGGCCCCCTGAGAAGCGGGTG 300
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         241 AGCTTGGAGGATTTCATGGAGAAGCTGGCACCCACCCACTGGCCCCCTGAGAAGCGGGTG 300

Qy         301 GCATACTGCTTTGAGGTGGCAGCCCAGCGAAGCCCAGATAAGAAGACGTGCCCCATGAAG 360
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         301 GCATACTGCTTTGAGGTGGCAGCCCAGCGAAGCCCAGATAAGAAGACGTGCCCCATGAAG 360

Qy         361 GAAGGAAACCCCTTTGGCCCATTCTGGGATCAGTTTCATGTGAGTTTCAACAAGTCGGAG 420
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         361 GAAGGAAACCCCTTTGGCCCATTCTGGGATCAGTTTCATGTGAGTTTCAACAAGTCGGAG 420

Qy         421 CTTTTTACAGGCATTTCCTTCAGTGCTTCCTACAGAGAACAATGGAGCCAGAGATTTTCT 480
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         421 CTTTTTACAGGCATTTCCTTCAGTGCTTCCTACAGAGAACAATGGAGCCAGAGATTTTCT 480

Qy         481 CCAAAGGAACATCCGGTGCTTGCCCTGCCAGGAGCCCCAGCCCAGTTCCCCGTCCTAGAA 540
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         481 CCAAAGGAACATCCGGTGCTTGCCCTGCCAGGAGCCCCAGCCCAGTTCCCCGTCCTAGAA 540

Qy         541 GAACACAGGCCACTACAGAAGTACATGGTATGGTCAGACGAAATGGTGAAGACGGGAGAG 600
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         541 GAACACAGGCCACTACAGAAGTACATGGTATGGTCAGACGAAATGGTGAAGACGGGAGAG 600

Qy         601 GCCCAGATTCATGCCCACCTTGTCCGGCCCTATGTGGGCATTCATCTGCGCATTGGCTCT 660
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         601 GCCCAGATTCATGCCCACCTTGTCCGGCCCTATGTGGGCATTCATCTGCGCATTGGCTCT 660

Qy         661 GACTGGAAGAACGCCTGTGCCATGCTGAAGGACGGGACTGCAGGCTCGCACTTCATGGCC 720
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         661 GACTGGAAGAACGCCTGTGCCATGCTGAAGGACGGGACTGCAGGCTCGCACTTCATGGCC 720

Qy         721 TCTCCGCAGTGTGTGGGCTACAGCCGCAGCACAGCGGCCCCCCTCACGATGACTATGTGC 780
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         721 TCTCCGCAGTGTGTGGGCTACAGCCGCAGCACAGCGGCCCCCCTCACGATGACTATGTGC 780

Qy         781 CTGCCTGACCTGAAGGAGATCCAGAGGGCTGTGAAGCTCTGGGTGAGGTCGCTGGATGCC 840
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         781 CTGCCTGACCTGAAGGAGATCCAGAGGGCTGTGAAGCTCTGGGTGAGGTCGCTGGATGCC 840

Qy         841 CAGTCGGTCTACGTTGCTACTGATTCCGAGAGTTATGTGCCTGAGCTCCAACAGCTCTTC 900
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         841 CAGTCGGTCTACGTTGCTACTGATTCCGAGAGTTATGTGCCTGAGCTCCAACAGCTCTTC 900

Qy         901 AAAGGGAAGGTGAAGGTGGTGAGCCTGAAGCCTGAGGTGGCCCAGGTCGACCTGTACATC 960
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         901 AAAGGGAAGGTGAAGGTGGTGAGCCTGAAGCCTGAGGTGGCCCAGGTCGACCTGTACATC 960

Qy         961 CTCGGCCAAGCCGACCACTTTATTGGCAACTGTGTCTCCTCCTTCACTGCCTTTGTGAAG 1020
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         961 CTCGGCCAAGCCGACCACTTTATTGGCAACTGTGTCTCCTCCTTCACTGCCTTTGTGAAG 1020

Qy        1021 CGGGAGCGGGACCTCCAGGGGAGGCCGTCTTCTTTCTTCGGCATGGACAGGCCCCCTAAG 1080
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1021 CGGGAGCGGGACCTCCAGGGGAGGCCGTCTTCTTTCTTCGGCATGGACAGGCCCCCTAAG 1080
```

```
Qy       1081 CTGCGGGACGAGTTCTGATT 1100
              ||||||||||||||||||||
Db       1081 CTGCGGGACGAGTTCTGATT 1100


RESULT 4
US-09-333-729A-2
; Sequence 2, Application US/09333729A
; Patent No. 6270987
; GENERAL INFORMATION:
;  APPLICANT:  Wang,Yang
;  APPLICANT:  Spellman,Michael W.
;  TITLE OF INVENTION:  O-Fucosyltransferase
;  FILE REFERENCE:  P1041P1D1-Substitute
;  CURRENT APPLICATION NUMBER: US/09/333,729A
;  CURRENT FILING DATE:  1999-06-15
;  PRIOR APPLICATION NUMBER:  US 08/798,741
;  PRIOR FILING DATE:  1997-11-26
;  NUMBER OF SEQ ID NOS:  21
; SEQ ID NO 2
;   LENGTH: 1514
;   TYPE: DNA
;   ORGANISM: Homo Sapien
US-09-333-729A-2

  Query Match           100.0%;  Score 1100;  DB 3;  Length 1514;
  Best Local Similarity  100.0%;  Pred. No. 0;
  Matches 1100;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 ATGCCCGCGGGCTCCTGGGACCCGGCCGGTTACCTGCTCTACTGCCCCTGCATGGGGCGC 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGCCCGCGGGCTCCTGGGACCCGGCCGGTTACCTGCTCTACTGCCCCTGCATGGGGCGC 60

Qy         61 TTTGGGAACCAGGCCGATCACTTCTTGGGCTCTCTGGCATTTGCAAAGCTGCTAAACCGT 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 TTTGGGAACCAGGCCGATCACTTCTTGGGCTCTCTGGCATTTGCAAAGCTGCTAAACCGT 120

Qy        121 ACCTTGGCTGTCCCTCCTTGGATTGAGTACCAGCATCACAAGCCTCCTTTCACCAACCTC 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 ACCTTGGCTGTCCCTCCTTGGATTGAGTACCAGCATCACAAGCCTCCTTTCACCAACCTC 180

Qy        181 CATGTGTCCTACCAGAAGTACTTCAAGCTGGAGCCCCTCCAGGCTTACCATCGGGTCATC 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 CATGTGTCCTACCAGAAGTACTTCAAGCTGGAGCCCCTCCAGGCTTACCATCGGGTCATC 240

Qy        241 AGCTTGGAGGATTTCATGGAGAAGCTGGCACCCACCCACTGGCCCCCTGAGAAGCGGGTG 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 AGCTTGGAGGATTTCATGGAGAAGCTGGCACCCACCCACTGGCCCCCTGAGAAGCGGGTG 300

Qy        301 GCATACTGCTTTGAGGTGGCAGCCCAGCGAAGCCCAGATAAGAAGACGTGCCCCATGAAG 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 GCATACTGCTTTGAGGTGGCAGCCCAGCGAAGCCCAGATAAGAAGACGTGCCCCATGAAG 360

Qy        361 GAAGGAAACCCCTTTGGCCCATTCTGGGATCAGTTTCATGTGAGTTTCAACAAGTCGGAG 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 GAAGGAAACCCCTTTGGCCCATTCTGGGATCAGTTTCATGTGAGTTTCAACAAGTCGGAG 420

Qy        421 CTTTTTACAGGCATTTCCTTCAGTGCTTCCTACAGAGAACAATGGAGCCAGAGATTTTCT 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 CTTTTTACAGGCATTTCCTTCAGTGCTTCCTACAGAGAACAATGGAGCCAGAGATTTTCT 480

Qy        481 CCAAAGGAACATCCGGTGCTTGCCCTGCCAGGAGCCCCAGCCCAGTTCCCCGTCCTAGAA 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 CCAAAGGAACATCCGGTGCTTGCCCTGCCAGGAGCCCCAGCCCAGTTCCCCGTCCTAGAA 540

Qy        541 GAACACAGGCCACTACAGAAGTACATGGTATGGTCAGACGAAATGGTGAAGACGGGAGAG 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 GAACACAGGCCACTACAGAAGTACATGGTATGGTCAGACGAAATGGTGAAGACGGGAGAG 600

Qy        601 GCCCAGATTCATGCCCACCTTGTCCGGCCCTATGTGGGCATTCATCTGCGCATTGGCTCT 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 GCCCAGATTCATGCCCACCTTGTCCGGCCCTATGTGGGCATTCATCTGCGCATTGGCTCT 660

Qy        661 GACTGGAAGAACGCCTGTGCCATGCTGAAGGACGGGACTGCAGGCTCGCACTTCATGGCC 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 GACTGGAAGAACGCCTGTGCCATGCTGAAGGACGGGACTGCAGGCTCGCACTTCATGGCC 720

Qy        721 TCTCCGCAGTGTGTGGGCTACAGCCGCAGCACAGCGGCCCCCCTCACGATGACTATGTGC 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 TCTCCGCAGTGTGTGGGCTACAGCCGCAGCACAGCGGCCCCCCTCACGATGACTATGTGC 780

Qy        781 CTGCCTGACCTGAAGGAGATCCAGAGGGCTGTGAAGCTCTGGGTGAGGTCGCTGGATGCC 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 CTGCCTGACCTGAAGGAGATCCAGAGGGCTGTGAAGCTCTGGGTGAGGTCGCTGGATGCC 840

Qy        841 CAGTCGGTCTACGTTGCTACTGATTCCGAGAGTTATGTGCCTGAGCTCCAACAGCTCTTC 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 CAGTCGGTCTACGTTGCTACTGATTCCGAGAGTTATGTGCCTGAGCTCCAACAGCTCTTC 900

Qy        901 AAAGGGAAGGTGAAGGTGGTGAGCCTGAAGCCTGAGGTGGCCCAGGTCGACCTGTACATC 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 AAAGGGAAGGTGAAGGTGGTGAGCCTGAAGCCTGAGGTGGCCCAGGTCGACCTGTACATC 960

Qy        961 CTCGGCCAAGCCGACCACTTTATTGGCAACTGTGTCTCCTCCTTCACTGCCTTTGTGAAG 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 CTCGGCCAAGCCGACCACTTTATTGGCAACTGTGTCTCCTCCTTCACTGCCTTTGTGAAG 1020

Qy       1021 CGGGAGCGGGACCTCCAGGGGAGGCCGTCTTCTTTCTTCGGCATGGACAGGCCCCCTAAG 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1021 CGGGAGCGGGACCTCCAGGGGAGGCCGTCTTCTTTCTTCGGCATGGACAGGCCCCCTAAG 1080

Qy       1081 CTGCGGGACGAGTTCTGATT 1100
              ||||||||||||||||||||
Db       1081 CTGCGGGACGAGTTCTGATT 1100


RESULT 5
US-08-978-741-4
; Sequence 4, Application US/08978741
; Patent No. 6100076
;  GENERAL INFORMATION:
;    APPLICANT:  Yang Wang, Michael W. Spellman
;    TITLE OF INVENTION:  O-Fucosyltransferase
;    NUMBER OF SEQUENCES:  17
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Genentech, Inc.
;      STREET:  1 DNA Way
;      CITY:  South San Francisco
;      STATE:  California
;      COUNTRY:  USA
;      ZIP:  94080
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  3.5 inch, 1.44 Mb floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  WinPatin (Genentech)
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/978,741
;      FILING DATE:  26-No. 6100076-1997
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/792498
;      FILING DATE:  31
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Svoboda, Craig G.
;      REGISTRATION NUMBER:  39,044
;      REFERENCE/DOCKET NUMBER:  P1041P1
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  650/225-1489
;      TELEFAX:  650/952-9881
;  INFORMATION FOR SEQ ID NO:  4:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1300 base pairs
;      TYPE:  Nucleic Acid
;      STRANDEDNESS:  Single
;      TOPOLOGY:  Linear
```

US-08-978-741-4

Query Match 99.9%; Score 1098.4; DB 3; Length 1300;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 1099; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```
Qy          1 ATGCCCGCGGGCTCCTGGGACCCGGCCGGTTACCTGCTCTACTGCCCCTGCATGGGGCGC 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        136 ATGCCCGCGGGCTCCTGGGACCCGGCCGGTTACCTGCTCTACTGCCCCTGCATGGGGCGC 195

Qy         61 TTTGGGAACCAGGCCGATCACTTCTTGGGCTCTCTGGCATTTGCAAAGCTGCTAAACCGT 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        196 TTTGGGAACCAGGCCGATCACTTCTTGGGCTCTCTGGCATTTGCAAAGCTGCTAAACCGT 255

Qy        121 ACCTTGGCTGTCCCTCCTTGGATTGAGTACCAGCATCACAAGCCTCCTTTCACCAACCTC 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        256 ACCTTGGCTGTCCCTCCTTGGATTGAGTACCAGCATCACAAGCCTCCTTTCACCAACCTC 315

Qy        181 CATGTGTCCTACCAGAAGTACTTCAAGCTGGAGCCCCTCCAGGCTTACCATCGGGTCATC 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        316 CATGTGTCCTACCAGAAGTACTTCAAGCTGGAGCCCCTCCAGGCTTACCATCGGGTCATC 375

Qy        241 AGCTTGGAGGATTTCATGGAGAAGCTGGCACCCACCCACTGGCCCCCTGAGAAGCGGGTG 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        376 AGCTTGGAGGATTTCATGGAGAAGCTGGCACCCACCCACTGGCCCCCTGAGAAGCGGGTG 435

Qy        301 GCATACTGCTTTGAGGTGGCAGCCCAGCGAAGCCCAGATAAGAAGACGTGCCCCATGAAG 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        436 GCATACTGCTTTGAGGTGGCAGCCCAGCGAAGCCCAGATAAGAAGACGTGCCCCATGAAG 495

Qy        361 GAAGGAAACCCCTTTGGCCCATTCTGGGATCAGTTTCATGTGAGTTTCAACAAGTCGGAG 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        496 GAAGGAAACCCCTTTGGCCCATTCTGGGATCAGTTTCATGTGAGTTTCAACAAGTCGGAG 555

Qy        421 CTTTTTACAGGCATTTCCTTCAGTGCTTCCTACAGAGAACAATGGAGCCAGAGATTTTCT 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        556 CTTTTTACAGGCATTTCCTTCAGTGCTTCCTACAGAGAACAATGGAGCCAGAGATTTTCT 615

Qy        481 CCAAAGGAACATCCGGTGCTTGCCCTGCCAGGAGCCCCAGCCCAGTTCCCCGTCCTAGAA 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        616 CCAAAGGAACATCCGGTGCTTGCCCTGCCAGGAGCCCCAGCCCAGTTCCCCGTCCTAGAG 675

Qy        541 GAACACAGGCCACTACAGAAGTACATGGTATGGTCAGACGAAATGGTGAAGACGGGAGAG 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        676 GAACACAGGCCACTACAGAAGTACATGGTATGGTCAGACGAAATGGTGAAGACGGGAGAG 735

Qy        601 GCCCAGATTCATGCCCACCTTGTCCGGCCCTATGTGGGCATTCATCTGCGCATTGGCTCT 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        736 GCCCAGATTCATGCCCACCTTGTCCGGCCCTATGTGGGCATTCATCTGCGCATTGGCTCT 795

Qy        661 GACTGGAAGAACGCCTGTGCCATGCTGAAGGACGGGACTGCAGGCTCGCACTTCATGGCC 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        796 GACTGGAAGAACGCCTGTGCCATGCTGAAGGACGGGACTGCAGGCTCGCACTTCATGGCC 855

Qy        721 TCTCCGCAGTGTGTGGGCTACAGCCGCAGCACAGCGGCCCCCCTCACGATGACTATGTGC 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        856 TCTCCGCAGTGTGTGGGCTACAGCCGCAGCACAGCGGCCCCCCTCACGATGACTATGTGC 915

Qy        781 CTGCCTGACCTGAAGGAGATCCAGAGGGCTGTGAAGCTCTGGGTGAGGTCGCTGGATGCC 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        916 CTGCCTGACCTGAAGGAGATCCAGAGGGCTGTGAAGCTCTGGGTGAGGTCGCTGGATGCC 975

Qy        841 CAGTCGGTCTACGTTGCTACTGATTCCGAGAGTTATGTGCCTGAGCTCCAACAGCTCTTC 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        976 CAGTCGGTCTACGTTGCTACTGATTCCGAGAGTTATGTGCCTGAGCTCCAACAGCTCTTC 1035

Qy        901 AAAGGGAAGGTGAAGGTGGTGAGCCTGAAGCCTGAGGTGGCCCAGGTCGACCTGTACATC 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1036 AAAGGGAAGGTGAAGGTGGTGAGCCTGAAGCCTGAGGTGGCCCAGGTCGACCTGTACATC 1095

Qy        961 CTCGGCCAAGCCGACCACTTTATTGGCAACTGTGTCTCCTCCTTCACTGCCTTTGTGAAG 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1096 CTCGGCCAAGCCGACCACTTTATTGGCAACTGTGTCTCCTCCTTCACTGCCTTTGTGAAG 1155

Qy       1021 CGGGAGCGGGACCTCCAGGGGAGGCCGTCTTCTTTCTTCGGCATGGACAGGCCCCCTAAG 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1156 CGGGAGCGGGACCTCCAGGGGAGGCCGTCTTCTTTCTTCGGCATGGACAGGCCCCCTAAG 1215

Qy       1081 CTGCGGGACGAGTTCTGATT 1100
              ||||||||||||||||||||
Db       1216 CTGCGGGACGAGTTCTGATT 1235
```

RESULT 6
US-09-333-729A-6
; Sequence 6, Application US/09333729A
; Patent No. 6270987
; GENERAL INFORMATION:
; APPLICANT: Wang,Yang
; APPLICANT: Spellman,Michael W.
; TITLE OF INVENTION: O-Fucosyltransferase
; FILE REFERENCE: P1041P1D1-Substitute
; CURRENT APPLICATION NUMBER: US/09/333,729A
; CURRENT FILING DATE: 1999-06-15
; PRIOR APPLICATION NUMBER: US 08/798,741
; PRIOR FILING DATE: 1997-11-26
; NUMBER OF SEQ ID NOS: 21
; SEQ ID NO 6
; LENGTH: 1300
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Plasmid insert.
US-09-333-729A-6

Query Match 99.9%; Score 1098.4; DB 3; Length 1300;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 1099; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```
Qy          1 ATGCCCGCGGGCTCCTGGGACCCGGCCGGTTACCTGCTCTACTGCCCCTGCATGGGGCGC 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        136 ATGCCCGCGGGCTCCTGGGACCCGGCCGGTTACCTGCTCTACTGCCCCTGCATGGGGCGC 195

Qy         61 TTTGGGAACCAGGCCGATCACTTCTTGGGCTCTCTGGCATTTGCAAAGCTGCTAAACCGT 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        196 TTTGGGAACCAGGCCGATCACTTCTTGGGCTCTCTGGCATTTGCAAAGCTGCTAAACCGT 255

Qy        121 ACCTTGGCTGTCCCTCCTTGGATTGAGTACCAGCATCACAAGCCTCCTTTCACCAACCTC 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        256 ACCTTGGCTGTCCCTCCTTGGATTGAGTACCAGCATCACAAGCCTCCTTTCACCAACCTC 315

Qy        181 CATGTGTCCTACCAGAAGTACTTCAAGCTGGAGCCCCTCCAGGCTTACCATCGGGTCATC 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        316 CATGTGTCCTACCAGAAGTACTTCAAGCTGGAGCCCCTCCAGGCTTACCATCGGGTCATC 375

Qy        241 AGCTTGGAGGATTTCATGGAGAAGCTGGCACCCACCCACTGGCCCCCTGAGAAGCGGGTG 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        376 AGCTTGGAGGATTTCATGGAGAAGCTGGCACCCACCCACTGGCCCCCTGAGAAGCGGGTG 435

Qy        301 GCATACTGCTTTGAGGTGGCAGCCCAGCGAAGCCCAGATAAGAAGACGTGCCCCATGAAG 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        436 GCATACTGCTTTGAGGTGGCAGCCCAGCGAAGCCCAGATAAGAAGACGTGCCCCATGAAG 495

Qy        361 GAAGGAAACCCCTTTGGCCCATTCTGGGATCAGTTTCATGTGAGTTTCAACAAGTCGGAG 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        496 GAAGGAAACCCCTTTGGCCCATTCTGGGATCAGTTTCATGTGAGTTTCAACAAGTCGGAG 555

Qy        421 CTTTTTACAGGCATTTCCTTCAGTGCTTCCTACAGAGAACAATGGAGCCAGAGATTTTCT 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        556 CTTTTTACAGGCATTTCCTTCAGTGCTTCCTACAGAGAACAATGGAGCCAGAGATTTTCT 615

Qy        481 CCAAAGGAACATCCGGTGCTTGCCCTGCCAGGAGCCCCAGCCCAGTTCCCCGTCCTAGAA 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        616 CCAAAGGAACATCCGGTGCTTGCCCTGCCAGGAGCCCCAGCCCAGTTCCCCGTCCTAGAG 675

Qy        541 GAACACAGGCCACTACAGAAGTACATGGTATGGTCAGACGAAATGGTGAAGACGGGAGAG 600
```

```
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        676 GAACACAGGCCACTACAGAAGTACATGGTATGGTCAGACGAAATGGTGAAGACGGGAGAG 735

Qy        601 GCCCAGATTCATGCCCACCTTGTCCGGCCCTATGTGGGCATTCATCTGCGCATTGGCTCT 660
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        736 GCCCAGATTCATGCCCACCTTGTCCGGCCCTATGTGGGCATTCATCTGCGCATTGGCTCT 795

Qy        661 GACTGGAAGAACGCCTGTGCCATGCTGAAGGACGGGACTGCAGGCTCGCACTTCATGGCC 720
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        796 GACTGGAAGAACGCCTGTGCCATGCTGAAGGACGGGACTGCAGGCTCGCACTTCATGGCC 855

Qy        721 TCTCCGCAGTGTGTGGGCTACAGCCGCAGCACAGCGGCCCCCCTCACGATGACTATGTGC 780
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        856 TCTCCGCAGTGTGTGGGCTACAGCCGCAGCACAGCGGCCCCCCTCACGATGACTATGTGC 915

Qy        781 CTGCCTGACCTGAAGGAGATCCAGAGGGCTGTGAAGCTCTGGGTGAGGTCGCTGGATGCC 840
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        916 CTGCCTGACCTGAAGGAGATCCAGAGGGCTGTGAAGCTCTGGGTGAGGTCGCTGGATGCC 975

Qy        841 CAGTCGGTCTACGTTGCTACTGATTCCGAGAGTTATGTGCCTGAGCTCCAACAGCTCTTC 900
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        976 CAGTCGGTCTACGTTGCTACTGATTCCGAGAGTTATGTGCCTGAGCTCCAACAGCTCTTC 1035

Qy        901 AAAGGGAAGGTGAAGGTGGTGAGCCTGAAGCCTGAGGTGGCCCAGGTCGACCTGTACATC 960
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1036 AAAGGGAAGGTGAAGGTGGTGAGCCTGAAGCCTGAGGTGGCCCAGGTCGACCTGTACATC 1095

Qy        961 CTCGGCCAAGCCGACCACTTTATTGGCAACTGTGTCTCCTCCTTCACTGCCTTTGTGAAG 1020
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1096 CTCGGCCAAGCCGACCACTTTATTGGCAACTGTGTCTCCTCCTTCACTGCCTTTGTGAAG 1155

Qy       1021 CGGGAGCGGGACCTCCAGGGGAGGCCGTCTTCTTTCTTCGGCATGGACAGGCCCCCTAAG 1080
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1156 CGGGAGCGGGACCTCCAGGGGAGGCCGTCTTCTTTCTTCGGCATGGACAGGCCCCCTAAG 1215

Qy       1081 CTGCGGGACGAGTTCTGATT 1100
                 ||||||||||||||||||||
Db       1216 CTGCGGGACGAGTTCTGATT 1235


RESULT 7
US-08-978-741-5
; Sequence 5, Application US/08978741
; Patent No. 6100076
;  GENERAL INFORMATION:
;    APPLICANT:  Yang Wang, Michael W. Spellman
;    TITLE OF INVENTION:  O-Fucosyltransferase
;    NUMBER OF SEQUENCES:  17
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Genentech, Inc.
;      STREET:  1 DNA Way
;      CITY:  South San Francisco
;      STATE:  California
;      COUNTRY:  USA
;      ZIP:  94080
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  3.5 inch, 1.44 Mb floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  WinPatin (Genentech)
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/978,741
;      FILING DATE:  26-No. 6100076-1997
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/792498
;      FILING DATE:  31
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Svoboda, Craig G.
;      REGISTRATION NUMBER:  39,044
;      REFERENCE/DOCKET NUMBER:  P1041P1
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  650/225-1489
;      TELEFAX:  650/952-9881
;  INFORMATION FOR SEQ ID NO:  5:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  11284 base pairs
;      TYPE:  Nucleic Acid
;      STRANDEDNESS:  Single
;      TOPOLOGY:  Linear
US-08-978-741-5

  Query Match             99.9%;  Score 1098.4;  DB 3;  Length 11284;
  Best Local Similarity   99.9%;  Pred. No. 0;
  Matches 1099;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 ATGCCCGCGGGCTCCTGGGACCCGGCCGGTTACCTGCTCTACTGCCCCTGCATGGGGCGC 60
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4236 ATGCCCGCGGGCTCCTGGGACCCGGCCGGTTACCTGCTCTACTGCCCCTGCATGGGGCGC 4295

Qy         61 TTTGGGAACCAGGCCGATCACTTCTTGGGCTCTCTGGCATTTGCAAAGCTGCTAAACCGT 120
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4296 TTTGGGAACCAGGCCGATCACTTCTTGGGCTCTCTGGCATTTGCAAAGCTGCTAAACCGT 4355

Qy        121 ACCTTGGCTGTCCCTCCTTGGATTGAGTACCAGCATCACAAGCCTCCTTTCACCAACCTC 180
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4356 ACCTTGGCTGTCCCTCCTTGGATTGAGTACCAGCATCACAAGCCTCCTTTCACCAACCTC 4415

Qy        181 CATGTGTCCTACCAGAAGTACTTCAAGCTGGAGCCCCTCCAGGCTTACCATCGGGTCATC 240
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4416 CATGTGTCCTACCAGAAGTACTTCAAGCTGGAGCCCCTCCAGGCTTACCATCGGGTCATC 4475

Qy        241 AGCTTGGAGGATTTCATGGAGAAGCTGGCACCCACCCACTGGCCCCCTGAGAAGCGGGTG 300
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4476 AGCTTGGAGGATTTCATGGAGAAGCTGGCACCCACCCACTGGCCCCCTGAGAAGCGGGTG 4535

Qy        301 GCATACTGCTTTGAGGTGGCAGCCCAGCGAAGCCCAGATAAGAAGACGTGCCCCATGAAG 360
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4536 GCATACTGCTTTGAGGTGGCAGCCCAGCGAAGCCCAGATAAGAAGACGTGCCCCATGAAG 4595

Qy        361 GAAGGAAACCCCCTTTGGCCCATTCTGGGATCAGTTTCATGTGAGTTTCAACAAGTCGGAG 420
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4596 GAAGGAAACCCCCTTTGGCCCATTCTGGGATCAGTTTCATGTGAGTTTCAACAAGTCGGAG 4655

Qy        421 CTTTTTACAGGCATTTCCTTCAGTGCTTCCTACAGAGAACAATGGAGCCAGAGATTTTCT 480
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4656 CTTTTTACAGGCATTTCCTTCAGTGCTTCCTACAGAGAACAATGGAGCCAGAGATTTTCT 4715

Qy        481 CCAAAGGAACATCCGGTGCTTGCCCTGCCAGGAGCCCCAGCCCAGTTCCCCGTCCTAGAA 540
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4716 CCAAAGGAACATCCGGTGCTTGCCCTGCCAGGAGCCCCAGCCCAGTTCCCCGTCCTAGAG 4775

Qy        541 GAACACAGGCCACTACAGAAGTACATGGTATGGTCAGACGAAATGGTGAAGACGGGAGAG 600
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4776 GAACACAGGCCACTACAGAAGTACATGGTATGGTCAGACGAAATGGTGAAGACGGGAGAG 4835

Qy        601 GCCCAGATTCATGCCCACCTTGTCCGGCCCTATGTGGGCATTCATCTGCGCATTGGCTCT 660
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4836 GCCCAGATTCATGCCCACCTTGTCCGGCCCTATGTGGGCATTCATCTGCGCATTGGCTCT 4895

Qy        661 GACTGGAAGAACGCCTGTGCCATGCTGAAGGACGGGACTGCAGGCTCGCACTTCATGGCC 720
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4896 GACTGGAAGAACGCCTGTGCCATGCTGAAGGACGGGACTGCAGGCTCGCACTTCATGGCC 4955

Qy        721 TCTCCGCAGTGTGTGGGCTACAGCCGCAGCACAGCGGCCCCCCTCACGATGACTATGTGC 780
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4956 TCTCCGCAGTGTGTGGGCTACAGCCGCAGCACAGCGGCCCCCCTCACGATGACTATGTGC 5015

Qy        781 CTGCCTGACCTGAAGGAGATCCAGAGGGCTGTGAAGCTCTGGGTGAGGTCGCTGGATGCC 840
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5016 CTGCCTGACCTGAAGGAGATCCAGAGGGCTGTGAAGCTCTGGGTGAGGTCGCTGGATGCC 5075

Qy        841 CAGTCGGTCTACGTTGCTACTGATTCCGAGAGTTATGTGCCTGAGCTCCAACAGCTCTTC 900
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5076 CAGTCGGTCTACGTTGCTACTGATTCCGAGAGTTATGTGCCTGAGCTCCAACAGCTCTTC 5135
```

```
Qy        901 AAAGGGAAGGTGAAGGTGGTGAGCCTGAAGCCTGAGGTGGCCCAGGTCGACCTGTACATC 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5136 AAAGGGAAGGTGAAGGTGGTGAGCCTGAAGCCTGAGGTGGCCCAGGTCGACCTGTACATC 5195

Qy        961 CTCGGCCAAGCCGACCACTTTATTGGCAACTGTGTCTCCTCCTTCACTGCCTTTGTGAAG 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5196 CTCGGCCAAGCCGACCACTTTATTGGCAACTGTGTCTCCTCCTTCACTGCCTTTGTGAAG 5255

Qy       1021 CGGGAGCGGGACCTCCAGGGGAGGCCGTCTTCTTTCTTCGGCATGGACAGGCCCCCTAAG 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5256 CGGGAGCGGGACCTCCAGGGGAGGCCGTCTTCTTTCTTCGGCATGGACAGGCCCCCTAAG 5315

Qy       1081 CTGCGGGACGAGTTCTGATT 1100
              ||||||||||||||||||||
Db       5316 CTGCGGGACGAGTTCTGATT 5335
```

RESULT 8
US-08-978-741-7
; Sequence 7, Application US/08978741
; Patent No. 6100076
; GENERAL INFORMATION:
; APPLICANT: Yang Wang, Michael W. Spellman
; TITLE OF INVENTION: O-Fucosyltransferase
; NUMBER OF SEQUENCES: 17
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genentech, Inc.
; STREET: 1 DNA Way
; CITY: South San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94080
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5 inch, 1.44 Mb floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: WinPatin (Genentech)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/978,741
; FILING DATE: 26-No. 6100076-1997
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/792498
; FILING DATE: 31
; ATTORNEY/AGENT INFORMATION:
; NAME: Svoboda, Craig G.
; REGISTRATION NUMBER: 39,044
; REFERENCE/DOCKET NUMBER: P1041P1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 650/225-1489
; TELEFAX: 650/952-9881
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 5009 base pairs
; TYPE: Nucleic Acid
; STRANDEDNESS: Single
; TOPOLOGY: Linear
US-08-978-741-7

Query Match 93.9%; Score 1033.4; DB 3; Length 5009;
Best Local Similarity 99.9%; Pred. No. 8.3e-302;
Matches 1034; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```
Qy          66 GAACCAGGCCGATCACTTCTTGGGCTCTCTGGCATTTGCAAAGCTGCTAAACCGTACCTT 125
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db           1 GAACCAGGCCGATCACTTCTTGGGCTCTCTGGCATTTGCAAAGCTGCTAAACCGTACCTT 60

Qy         126 GGCTGTCCCTCCTTGGATTGAGTACCAGCATCACAAGCCTCCTTTCACCAACCTCCATGT 185
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          61 GGCTGTCCCTCCTTGGATTGAGTACCAGCATCACAAGCCTCCTTTCACCAACCTCCATGT 120

Qy         186 GTCCTACCAGAAGTACTTCAAGCTGGAGCCCCTCCAGGCTTACCATCGGGTCATCAGCTT 245
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         121 GTCCTACCAGAAGTACTTCAAGCTGGAGCCCCTCCAGGCTTACCATCGGGTCATCAGCTT 180

Qy         246 GGAGGATTTCATGGAGAAGCTGGCACCCACCCACTGGCCCCCTGAGAAGCGGGTGGCATA 305
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         181 GGAGGATTTCATGGAGAAGCTGGCACCCACCCACTGGCCCCCTGAGAAGCGGGTGGCATA 240

Qy         306 CTGCTTTGAGGTGGCAGCCCAGCGAAGCCCAGATAAGAAGACGTGCCCCATGAAGGAAGG 365
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         241 CTGCTTTGAGGTGGCAGCCCAGCGAAGCCCAGATAAGAAGACGTGCCCCATGAAGGAAGG 300

Qy         366 AAACCCCTTTGGCCCATTCTGGGATCAGTTTCATGTGAGTTTCAACAAGTCGGAGCTTTT 425
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         301 AAACCCCTTTGGCCCATTCTGGGATCAGTTTCATGTGAGTTTCAACAAGTCGGAGCTTTT 360

Qy         426 TACAGGCATTTCCTTCAGTGCTTCCTACAGAGAACAATGGAGCCAGAGATTTTCTCCAAA 485
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         361 TACAGGCATTTCCTTCAGTGCTTCCTACAGAGAACAATGGAGCCAGAGATTTTCTCCAAA 420

Qy         486 GGAACATCCGGTGCTTGCCCTGCCAGGAGCCCCAGCCCAGTTCCCCGTCCTAGAAGAACA 545
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||| |||||
Db         421 GGAACATCCGGTGCTTGCCCTGCCAGGAGCCCCAGCCCAGTTCCCCGTCCTAGAGGAACA 480

Qy         546 CAGGCCACTACAGAAGTACATGGTATGGTCAGACGAAATGGTGAAGACGGGAGAGGCCCA 605
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         481 CAGGCCACTACAGAAGTACATGGTATGGTCAGACGAAATGGTGAAGACGGGAGAGGCCCA 540

Qy         606 GATTCATGCCCACCTTGTCCGGCCCTATGTGGGCATTCATCTGCGCATTGGCTCTGACTG 665
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         541 GATTCATGCCCACCTTGTCCGGCCCTATGTGGGCATTCATCTGCGCATTGGCTCTGACTG 600

Qy         666 GAAGAACGCCTGTGCCATGCTGAAGGACGGGACTGCAGGCTCGCACTTCATGGCCTCTCC 725
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         601 GAAGAACGCCTGTGCCATGCTGAAGGACGGGACTGCAGGCTCGCACTTCATGGCCTCTCC 660

Qy         726 GCAGTGTGTGGGCTACAGCCGCAGCACAGCGGCCCCCCTCACGATGACTATGTGCCTGCC 785
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         661 GCAGTGTGTGGGCTACAGCCGCAGCACAGCGGCCCCCCTCACGATGACTATGTGCCTGCC 720

Qy         786 TGACCTGAAGGAGATCCAGAGGGCTGTGAAGCTCTGGGTGAGGTCGCTGGATGCCCAGTC 845
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         721 TGACCTGAAGGAGATCCAGAGGGCTGTGAAGCTCTGGGTGAGGTCGCTGGATGCCCAGTC 780

Qy         846 GGTCTACGTTGCTACTGATTCCGAGAGTTATGTGCCTGAGCTCCAACAGCTCTTCAAAGG 905
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         781 GGTCTACGTTGCTACTGATTCCGAGAGTTATGTGCCTGAGCTCCAACAGCTCTTCAAAGG 840

Qy         906 GAAGGTGAAGGTGGTGAGCCTGAAGCCTGAGGTGGCCCAGGTCGACCTGTACATCCTCGG 965
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         841 GAAGGTGAAGGTGGTGAGCCTGAAGCCTGAGGTGGCCCAGGTCGACCTGTACATCCTCGG 900

Qy         966 CCAAGCCGACCACTTTATTGGCAACTGTGTCTCCTCCTTCACTGCCTTTGTGAAGCGGGA 1025
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         901 CCAAGCCGACCACTTTATTGGCAACTGTGTCTCCTCCTTCACTGCCTTTGTGAAGCGGGA 960

Qy        1026 GCGGGACCTCCAGGGGAGGCCGTCTTCTTTCTTCGGCATGGACAGGCCCCCTAAGCTGCG 1085
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         961 GCGGGACCTCCAGGGGAGGCCGTCTTCTTTCTTCGGCATGGACAGGCCCCCTAAGCTGCG 1020

Qy        1086 GGACGAGTTCTGATT 1100
               ||||||||||||||||
Db        1021 GGACGAGTTCTGATT 1035
```

RESULT 9
US-09-333-729A-8
; Sequence 8, Application US/09333729A
; Patent No. 6270987
; GENERAL INFORMATION:
; APPLICANT: Wang,Yang
; APPLICANT: Spellman,Michael W.
; TITLE OF INVENTION: O-Fucosyltransferase
; FILE REFERENCE: P1041P1D1-Substitute
; CURRENT APPLICATION NUMBER: US/09/333,729A

```
;  CURRENT FILING DATE:  1999-06-15
;  PRIOR APPLICATION NUMBER:  US 08/798,741
;  PRIOR FILING DATE:  1997-11-26
;  NUMBER OF SEQ ID NOS:  21
; SEQ ID NO 8
;   LENGTH: 5009
;   TYPE: DNA
;   ORGANISM: Homo Sapien
US-09-333-729A-8

  Query Match             93.9%;  Score 1033.4;  DB 3;  Length 5009;
  Best Local Similarity   99.9%;  Pred. No. 8.3e-302;
  Matches 1034;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy         66 GAACCAGGCCGATCACTTCTTGGGCTCTCTGGCATTTGCAAAGCTGCTAAACCGTACCTT 125
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 GAACCAGGCCGATCACTTCTTGGGCTCTCTGGCATTTGCAAAGCTGCTAAACCGTACCTT 60

Qy        126 GGCTGTCCCTCCTTGGATTGAGTACCAGCATCACAAGCCTCCTTTCACCAACCTCCATGT 185
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GGCTGTCCCTCCTTGGATTGAGTACCAGCATCACAAGCCTCCTTTCACCAACCTCCATGT 120

Qy        186 GTCCTACCAGAAGTACTTCAAGCTGGAGCCCCTCCAGGCTTACCATCGGGTCATCAGCTT 245
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 GTCCTACCAGAAGTACTTCAAGCTGGAGCCCCTCCAGGCTTACCATCGGGTCATCAGCTT 180

Qy        246 GGAGGATTTCATGGAGAAGCTGGCACCCACCCACTGGCCCCCTGAGAAGCGGGTGGCATA 305
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 GGAGGATTTCATGGAGAAGCTGGCACCCACCCACTGGCCCCCTGAGAAGCGGGTGGCATA 240

Qy        306 CTGCTTTGAGGTGGCAGCCCAGCGAAGCCCAGATAAGAAGACGTGCCCCATGAAGGAAGG 365
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 CTGCTTTGAGGTGGCAGCCCAGCGAAGCCCAGATAAGAAGACGTGCCCCATGAAGGAAGG 300

Qy        366 AAACCCCTTTGGCCCATTCTGGGATCAGTTTCATGTGAGTTTCAACAAGTCGGAGCTTTT 425
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 AAACCCCTTTGGCCCATTCTGGGATCAGTTTCATGTGAGTTTCAACAAGTCGGAGCTTTT 360

Qy        426 TACAGGCATTTCCTTCAGTGCTTCCTACAGAGAACAATGGAGCCAGAGATTTTCTCCAAA 485
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 TACAGGCATTTCCTTCAGTGCTTCCTACAGAGAACAATGGAGCCAGAGATTTTCTCCAAA 420

Qy        486 GGAACATCCGGTGCTTGCCCTGCCAGGAGCCCCAGCCCAGTTCCCCGTCCTAGAAGAACA 545
              |||||||||||||||||||||||||||||||||||||||||||||||||||||| |||||
Db        421 GGAACATCCGGTGCTTGCCCTGCCAGGAGCCCCAGCCCAGTTCCCCGTCCTAGAGGAACA 480

Qy        546 CAGGCCACTACAGAAGTACATGGTATGGTCAGACGAAATGGTGAAGACGGGAGAGGCCCA 605
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 CAGGCCACTACAGAAGTACATGGTATGGTCAGACGAAATGGTGAAGACGGGAGAGGCCCA 540

Qy        606 GATTCATGCCCACCTTGTCCGGCCCTATGTGGGCATTCATCTGCGCATTGGCTCTGACTG 665
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 GATTCATGCCCACCTTGTCCGGCCCTATGTGGGCATTCATCTGCGCATTGGCTCTGACTG 600

Qy        666 GAAGAACGCCTGTGCCATGCTGAAGGACGGGACTGCAGGCTCGCACTTCATGGCCTCTCC 725
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 GAAGAACGCCTGTGCCATGCTGAAGGACGGGACTGCAGGCTCGCACTTCATGGCCTCTCC 660

Qy        726 GCAGTGTGTGGGCTACAGCCGCAGCACAGCGGCCCCCCTCACGATGACTATGTGCCTGCC 785
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 GCAGTGTGTGGGCTACAGCCGCAGCACAGCGGCCCCCCTCACGATGACTATGTGCCTGCC 720

Qy        786 TGACCTGAAGGAGATCCAGAGGGCTGTGAAGCTCTGGGTGAGGTCGCTGGATGCCCAGTC 845
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 TGACCTGAAGGAGATCCAGAGGGCTGTGAAGCTCTGGGTGAGGTCGCTGGATGCCCAGTC 780

Qy        846 GGTCTACGTTGCTACTGATTCCGAGAGTTATGTGCCTGAGCTCCAACAGCTCTTCAAAGG 905
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 GGTCTACGTTGCTACTGATTCCGAGAGTTATGTGCCTGAGCTCCAACAGCTCTTCAAAGG 840

Qy        906 GAAGGTGAAGGTGGTGAGCCTGAAGCCTGAGGTGGCCCAGGTCGACCTGTACATCCTCGG 965
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 GAAGGTGAAGGTGGTGAGCCTGAAGCCTGAGGTGGCCCAGGTCGACCTGTACATCCTCGG 900

Qy        966 CCAAGCCGACCACTTTATTGGCAACTGTGTCTCCTCCTTCACTGCCTTTGTGAAGCGGGA 1025
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 CCAAGCCGACCACTTTATTGGCAACTGTGTCTCCTCCTTCACTGCCTTTGTGAAGCGGGA 960

Qy       1026 GCGGGACCTCCAGGGGAGGCCGTCTTCTTTCTTCGGCATGGACAGGCCCCCTAAGCTGCG 1085
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 GCGGGACCTCCAGGGGAGGCCGTCTTCTTTCTTCGGCATGGACAGGCCCCCTAAGCTGCG 1020

Qy       1086 GGACGAGTTCTGATT 1100
              |||||||||||||||
Db       1021 GGACGAGTTCTGATT 1035


RESULT 10
US-09-270-767-14353
; Sequence 14353, Application US/09270767
; Patent No. 6703491
; GENERAL INFORMATION:
;  APPLICANT: Homburger et al.
;  TITLE OF INVENTION: Nucleic acids and proteins of Drosophila melanogaster
;  FILE REFERENCE: File Reference: 7326-094
;  CURRENT APPLICATION NUMBER: US/09/270,767
;  CURRENT FILING DATE: 1999-03-17
;  NUMBER OF SEQ ID NOS: 62517
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 14353
;   LENGTH: 1320
;   TYPE: DNA
;   ORGANISM: Drosophila melanogaster
US-09-270-767-14353

  Query Match             15.3%;  Score 168;  DB 4;  Length 1320;
  Best Local Similarity   51.6%;  Pred. No. 2.9e-40;
  Matches  564;  Conservative    0;  Mismatches  485;  Indels   45;  Gaps    6;

Qy         17 GGGACCCGGCCGGTTACCTGCTCTACTGCCCCTGCATGGGGCGCTTTGGGAACCAGGCCG 76
              | || ||    || |||||   |||||| || || ||||| |||||||| ||||||||||
Db         73 GCGATCCCAATGGCTACCTCACCTACTGTCCGTGTATGGGACGCTTTGGCAACCAGGCCG 132

Qy         77 ATCACTTCTTGGGCTCTCTGGCATTTGCAAAGCTGCTAAACCGTACCTTGGCTGTCCCTC 136
              | |||||| |||| ||  |||| || || |||  ||| || || ||| ||    | || |
Db        133 ACCACTTCCTGGGATCATTGGCCTTCGCCAAGGCGCTTAATCGCACCCTGATCCTGCCGC 192

Qy        137 CTTGGATTGAGTACCAGCATCACAAGCCTCCTTTCACCAACCTCCATGTGTCCTACCAGA 196
              | ||| | ||||      |||  | |  |    |      ||  || ||  | | | | | |
Db        193 CGTGGGTGGAGT------ATCGTAGGGGTGAACTGCGATCCCGTCAGGTACCGTTCAACA 246

Qy        197 AGTACTTCAAGCTGGAGCCCCTCCAGGCTTACCATCGGGTCATCAGCTTGGAGGATTTCA 256
                || ||  |  ||||||||||  |||  ||||||||| ||||||| | |||  ||||||||
Db        247 CATATTTTGAAGTGGAGCCCCTGAAGGAATACCATCGCGTCATCACCATGGCAGATTTCA 306

Qy        257 TGGAGAAGCTGGCACCCACCCACTGGCCCCCTGAGAAGCGGGTGGCATACTGCTTTGAGG 316
              ||  | | ||||||   |  |   |||||||      | |||| ||| |||  |||||   |||
Db        307 TGTGGCACCTGGCCGACGACATTTGGCCAGAATCGGAGCGAGTGTCATTTTGCTACAAGG 366

Qy        317 ---------------TGGCAGCCCAGCGAAGCCCAGATAAGAAGACGTGCCCCATGAAGG 361
                              |   |   ||      ||||| |||   |  |||| |   ||||
Db        367 AACGATATAGCCTTCAGCAGGAGAAGAACGATCCAGACAAGCCCAATTGCCACGCCAAGG 426

Qy        362 AAGGAAACCCCTTTGGCCCATTCTGGGATCAGTTTCATGTGAGTTTCAACAAGTCGGAGC 421
              | || || || |||||| || || |||||    |||||  |    ||        ||| ||| 
Db        427 ATGGCAATCCTTTTGGTCCCTTTTGGGACACTTTTCACATAGACTTTGTGCGGTCAGAGT 486

Qy        422 TTTTTACAGGCATTTCCTTCAGTGCTTCCTACAG---------AGAACAATGGAGCCAGA 472
              | | | |     ||   ||   ||        | ||          |   | |||      |
Db        487 TCTATGCGCCACTTCATTTTGATGTGCATCATAGCAACGAGGCTGCCAAGTGGCAGACCA 546

Qy        473 GATTTTCTCCAAAGGAACATCCGGTGCTTGCCCTGCCAGGAGCCCCAGCCCAGTTCCCCG 532
               || | || || | |  | |||| || || |  |  | || ||| || || |   || || |
Db        547 AATATCCTGCAGAATCATATCCCGTACTCGCGTTCACCGGAGCTCCGGCTAGTTTTCCTG 606

Qy        533 TCCTAGAAGAACACAGGCCACTACAGAAGTACATGGTATGGTCAGACGAAATGGTGAAGA 592
```

```
Db        607 TTCAGCTAGAGAACTGCAAGCTGCAGCGCTACTTGCAGTGGAGTCAACGGTATAGGGAAG 666

Qy        593 CGGGAGAGGCCCAGATTCATGCCCACCTTGTCCGGC---CCTATGTGGGCATTCATCTGC 649

Db        667 CATCTAAGGATTTCATCCGAGAGCAGTTGCCTCGGGGTGCCTTTTTGGGCATTCATCTGC 726

Qy        650 GCATTGGCTCTGACTGGAAGAACGCCTGTGCCATGCTGAAGGACGG---------GACTG 700

Db        727 GCAACGGTATCGATTGGGTGAGAGCCTGTGAGCACGTCAAGGATAGCCAGCATCTGTTTG 786

Qy        701 CAGGCTCGCACTTCATGGCCTCTCCGCAGTGTGTGGGCTACAGCCGCAGCACAGCGGCCC 760

Db        787 CCTCGCCGCAGTGCTTGGGCTATAAAAATGAACGTGGTGCACTCTACCCGGAGCTCTGCA 846

Qy        761 CCCTCACGATGACTATGTGCCTGCCTGACCTGAAGGAGATCCAGAGGGCTGTGAAGCTCT 820

Db        847 TGCCCTCCAAGGAGGCGATCATCCGCCAGCTAAAGAGAACCATTAAGAACGTGCGCCAAA 906

Qy        821 GGGTGAGGTCGCTGGATGCCCAGTCGGTCTACGTTGCTACTGATTCCGAGAGTTATGTGC 880

Db        907 CTCAGCCGGACAACGAAATCAAATCAGTTTTCGTGGCGTCAGACTCCAATCACATGATTG 966

Qy        881 CTGAGCTCCAACAGCTCTTCAAAGGGAAGG---TGAAGGTGGTGAGCCTGAAGCCTGAGG 937

Db        967 GTGAACTAAACACGGCCCTTAGTCGCATGGGCATCAGTGTGCACAAGCTGACGGAGGATG 1026

Qy        938 TGGCCCAGGTCGACCTGTACATCCTCGGCCAAGCCGACCACTTTATTGGCAACTGTGTCT 997

Db       1027 ATCCTTACCTGGACTTGGCCATTCTCGGACAGTCGAACCACTTTATCGGCAACTGTATAT 1086

Qy        998 CCTCCTTCACTGCCTTTGTGAAGCGGGAGCGGGACCTCCAGGGGAGGCCGTCTTCTTTCT 1057

Db       1087 CCTCTTACTCGGCATTCGTAAAAAGGGAACGAGATGTGCACGGTTTTCCATCGTACTTCT 1146

Qy       1058 TCGGCATGGACAGG 1071

Db       1147 GGGGATTCCCCAAG 1160


RESULT 11
US-09-270-767-30507
; Sequence 30507, Application US/09270767
; Patent No. 6703491
; GENERAL INFORMATION:
;  APPLICANT: Homburger et al.
;  TITLE OF INVENTION: Nucleic acids and proteins of Drosophila melanogaster
;  FILE REFERENCE: File Reference: 7326-094
;  CURRENT APPLICATION NUMBER: US/09/270,767
;  CURRENT FILING DATE: 1999-03-17
;  NUMBER OF SEQ ID NOS: 62517
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 30507
;   LENGTH: 722
;   TYPE: DNA
;   ORGANISM: Drosophila melanogaster
US-09-270-767-30507

  Query Match             5.7%;  Score 62.2;  DB 4;  Length 722;
  Best Local Similarity   49.5%;  Pred. No. 2.2e-08;
  Matches  225;  Conservative    0;  Mismatches  218;  Indels   12;  Gaps    2;

Qy        629 CCTATGTGGGCATTCATCTGCGCATTGGCTCTGACTGGAAGAACGCCTGTGCCATGCTGA 688

Db        108 CCTTTTTGGGCATTCATCTGCGCAACGGTATCGATTGGGTGAGAGCCTGTGAGCACGTCA 167

Qy        689 AGGACGG---------GACTGCAGGCTCGCACTTCATGGCCTCTCCGCAGTGTGTGGGCT 739

Db        168 AGGATAGCCAGCATCTGTTTGCCTCGCCGCAGTGCTTGGGCTATAAAAATGAACGTGGTG 227

Qy        740 ACAGCCGCAGCACAGCGGCCCCCCTCACGATGACTATGTGCCTGCCTGACCTGAAGGAGA 799

Db        228 CACTCTACCCGGAGCTCTGCATGCCCTCCAAGGAGGCGATCATCCGCCAGCTAAAGAGAA 287

Qy        800 TCCAGAGGGCTGTGAAGCTCTGGGTGAGGTCGCTGGATGCCCAGTCGGTCTACGTTGCTA 859

Db        288 CCATTAAGAACGTGCGCCAAACTCAGCCGGACAACGAAATCAAATCAGTTTTCGTGGCGT 347

Qy        860 CTGATTCCGAGAGTTATGTGCCTGAGCTCCAACAGCTCTTCAAAGGGAAGG---TGAAGG 916

Db        348 CAGACTCCAATCACATGATTGGTGAACTAAACACGGCCCTTAGTCGCATGGGCATCAGTG 407

Qy        917 TGGTGAGCCTGAAGCCTGAGGTGGCCCAGGTCGACCTGTACATCCTCGGCCAAGCCGACC 976

Db        408 TGCACAAGCTGACGGAGGATGATCCTTACCTGGACTTGGCCATTCTCGGACAGTCGAACC 467

Qy        977 ACTTTATTGGCAACTGTGTCTCCTCCTTCACTGCCTTTGTGAAGCGGGAGCGGGACCTCC 1036

Db        468 ACTTTATCGGCAACTGTATATCCTCTTACTCGGCATTCGTAAAAAGGGAACGAGATGTGC 527

Qy       1037 AGGGGAGGCCGTCTTCTTTCTTCGGCATGGACAGG 1071

Db        528 ACGGTTTTCCATCGTACTTCTGGGGATTCCCCAAG 562


RESULT 12
US-09-270-767-28905/c
; Sequence 28905, Application US/09270767
; Patent No. 6703491
; GENERAL INFORMATION:
;  APPLICANT: Homburger et al.
;  TITLE OF INVENTION: Nucleic acids and proteins of Drosophila melanogaster
;  FILE REFERENCE: File Reference: 7326-094
;  CURRENT APPLICATION NUMBER: US/09/270,767
;  CURRENT FILING DATE: 1999-03-17
;  NUMBER OF SEQ ID NOS: 62517
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 28905
;   LENGTH: 998
;   TYPE: DNA
;   ORGANISM: Drosophila melanogaster
US-09-270-767-28905

  Query Match             5.3%;  Score 58.6;  DB 4;  Length 998;
  Best Local Similarity   60.2%;  Pred. No. 3.2e-07;
  Matches   97;  Conservative    0;  Mismatches   64;  Indels    0;  Gaps    0;

Qy        911 TGAAGGTGGTGAGCCTGAAGCCTGAGGTGGCCCAGGTCGACCTGTACATCCTCGGCCAAG 970

Db        896 TCAGTGTGCACAAGCTGACGGAGGATGATCCTTACCTGGACTTGGCCATTCTCGGACAGT 837

Qy        971 CCGACCACTTTATTGGCAACTGTGTCTCCTCCTTCACTGCCTTTGTGAAGCGGGAGCGGG 1030

Db        836 CGAACCACTTTATCGGCAACTGTATATCCTCTTACTCGGCATTCGTAAAAAGGGAACGAG 777

Qy       1031 ACCTCCAGGGGAGGCCGTCTTCTTTCTTCGGCATGGACAGG 1071

Db        776 ATGTGCACGGTTTTCCATCGTACTTCTGGGGATTCCCCAAG 736


RESULT 13
US-09-270-767-13022/c
; Sequence 13022, Application US/09270767
; Patent No. 6703491
; GENERAL INFORMATION:
;  APPLICANT: Homburger et al.
;  TITLE OF INVENTION: Nucleic acids and proteins of Drosophila melanogaster
;  FILE REFERENCE: File Reference: 7326-094
;  CURRENT APPLICATION NUMBER: US/09/270,767
;  CURRENT FILING DATE: 1999-03-17
;  NUMBER OF SEQ ID NOS: 62517
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 13022
;   LENGTH: 2034
;   TYPE: DNA
;   ORGANISM: Drosophila melanogaster
```

US-09-270-767-13022

Query Match 5.3%; Score 58.6; DB 4; Length 2034;
Best Local Similarity 60.2%; Pred. No. 4.8e-07;
Matches 97; Conservative 0; Mismatches 64; Indels 0; Gaps 0;

```
Qy        911 TGAAGGTGGTGAGCCTGAAGCCTGAGGTGGCCCAGGTCGACCTGTACATCCTCGGCCAAG 970
              | |  |||   |  |||| |   || |   |  |  | ||| ||  ||| ||||| ||
Db       1932 TCAGTGTGCACAAGCTGACGGAGGATGATCCTTACCTGGACTTGGCCATTCTCGGACAGT 1873

Qy        971 CCGACCACTTTATTGGCAACTGTGTCTCCTCCTTCACTGCCTTTGTGAAGCGGGAGCGGG 1030
              |  |||||||||| |||||||| | |||| | | | || || || ||  |||| || |
Db       1872 CGAACCACTTTATCGGCAACTGTATATCCTCTTACTCGGCATTCGTAAAAAGGGAACGAG 1813

Qy       1031 ACCTCCAGGGGAGGCCGTCTTCTTTCTTCGGCATGGACAGG 1071
              |  | || ||    || || |  ||||  ||  |   || |
Db       1812 ATGTGCACGGTTTTCCATCGTACTTCTGGGGATTCCCCAAG 1772
```

RESULT 14
US-09-949-016-12127/c
; Sequence 12127, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 12127
; LENGTH: 19227
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-12127

Query Match 5.1%; Score 56; DB 4; Length 19227;
Best Local Similarity 100.0%; Pred. No. 1.1e-05;
Matches 56; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 ATGCCCGCGGGCTCCTGGGACCCGGCCGGTTACCTGCTCTACTGCCCCTGCATGGG 56
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1718 ATGCCCGCGGGCTCCTGGGACCCGGCCGGTTACCTGCTCTACTGCCCCTGCATGGG 1663
```

RESULT 15
US-09-949-016-16285/c
; Sequence 16285, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 16285
; LENGTH: 19228
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-16285

Query Match 5.1%; Score 56; DB 4; Length 19228;
Best Local Similarity 100.0%; Pred. No. 1.1e-05;
Matches 56; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 ATGCCCGCGGGCTCCTGGGACCCGGCCGGTTACCTGCTCTACTGCCCCTGCATGGG 56
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1718 ATGCCCGCGGGCTCCTGGGACCCGGCCGGTTACCTGCTCTACTGCCCCTGCATGGG 1663
```

Search completed: November 9, 2005, 10:46:08
Job time : 202.044 secs